

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

DIPLOMOVÁ PRÁCE



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

**FAKULTA ELEKTROTECHNIKY
A KOMUNIKAČNÍCH TECHNOLOGIÍ**

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV TELEKOMUNIKACÍ

DEPARTMENT OF TELECOMMUNICATIONS

**URČOVÁNÍ OBDOBÍ VZNIKU INTERPRETACE ZA POMOCI
METOD PARAMETRIZACE HUDEBNÍHO SIGNÁLU**

RECOGNIZING THE HISTORICAL PERIOD OF INTERPRETATION BASED ON THE MUSIC SIGNAL
PARAMETERIZATION

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

Bc. Vítězslav Král

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Tomáš Kiska

BRNO 2018

Diplomová práce

magisterský navazující studijní obor **Audio inženýrství**

Ústav telekomunikací

Student: Bc. Vítězslav Král

ID: 164599

Ročník: 2

Akademický rok: 2017/18

NÁZEV TÉMATU:

Určování období vzniku interpretace za pomoci metod parametrizace hudebního signálu

POKYNY PRO VYPRACOVÁNÍ:

Shrňte dosavadní poznatky z oblasti srovnávání zvukových nahrávek v oblasti tzv. „Music information retrieval“. Dále analyzujte interpretace z hlediska tempa, dynamiky a barvy zvuku. Implementujte systém, který za pomoci metod strojového učení a korelační analýzy vybere takové hudební parametry, které mají největší schopnost diferencovat období vzniku jednotlivých interpretací.

DOPORUČENÁ LITERATURA:

[1] SCHEDL, M.; GOMÉZ, E; URBANO, J. Music Information Retrieval: Recent Developments and Applications. Foundations and Trends in Information Retrieval, 2014: s. 127-261, DOI: 978-1-60198-807-2.

[2] MÜLLER, M. Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications [online]. Springer International Publishing Switzerland, 2015, 483 s. ISBN 978-3-319-21945-5.

Termín zadání: 5.2.2018

Termín odevzdání: 21.5.2018

Vedoucí práce: Ing. Tomáš Kiska

Konzultant:

prof. Ing. Jiří Mišurec, CSc.
předseda oborové rady

UPOZORNĚNÍ:

Autor diplomové práce nesmí při vytváření diplomové práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

ABSTRAKT

Cílem této diplomové práce je shrnout dosavadní poznatky z oblasti srovnávání zvukových nahrávek a implementovat vyhodnocovací systém pro určení období vzniku za pomoci metod parametrizace hudebního signálu. V první části této práce jsou popsány reprezentace, jakých hudba může nabývat. Dále je uveden průřez parametrů, které mohou být z hudebních nahrávek extrahovány a poskytují informaci o dynamice, tempu, barvě či časovém vývoji hudební nahrávky. V části druhé je popsán vyhodnocovací systém a jeho jednotlivé dílčí bloky. Vstupními daty pro tento vyhodnocovací systém je vytvořená databáze, čítající 56 zvukových nahrávek první věty Beethovenovy páté symfonie. Poslední kapitola je věnována shrnutí dosažených výsledků.

KLÍČOVÁ SLOVA

klasifikace, parametry hudebních nahrávek, selekce parametrů, SFFS, statistická analýza, získávání informací z hudby

ABSTRACT

The aim of this semestral work is to summarize the existing knowledge from the area of comparison of musical recordings and to implement an evaluation system for determining the period of creation using the music signal parameterization. In the first part of this work are describe representations which can music take. Next, there is a cross-section of parameters that can be extracted from music recordings provides information on the dynamics, tempo, color, or time development of the music's recording. In the second part is described evaluation system and its individual sub-blocks. The input data for this evaluation system is a database of 56 sound recordings of the first movement of Beethoven's 5th Symphony. The last chapter is dedicated to a summary of the achieved results.

KEYWORDS

classification, features selection, musical features, music information retrieval, SFFS, statistical analysis

KRÁL, Vítězslav. *Určování období vzniku interpretace za pomoci metod parametrizace hudebního signálu*. Brno, 2017, 72 s. Diplomová práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací. Vedoucí práce: Ing. Tomáš Kiska

PROHLÁŠENÍ

Prohlašuji, že svou diplomovou práci na téma „Určování období vzniku interpretace za pomoci metod parametrizace hudebního signálu“ jsem vypracoval samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené diplomové práce dále prohlašuji, že v souvislosti s vytvořením této diplomové práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

Brno

.....

podpis autora

PODĚKOVÁNÍ

Rád bych poděkoval vedoucímu diplomové práce panu Ing. Tomáši Kiskovi za odborné vedení, konzultace, trpělivost a podnětné návrhy k práci.

Brno

.....

podpis autora

PODĚKOVÁNÍ

Výzkum popsáný v této diplomové práci byl realizován v laboratořích podpořených z projektu SIX; registrační číslo CZ.1.05/2.1.00/03.0072, operační program Výzkum a vývoj pro inovace.

Brno

.....
podpis autora

OBSAH

Úvod	12
1 Získávání informací z hudebních nahrávek	13
1.1 Motivace	13
1.2 Historie MIR	13
2 Reprezentace hudby	14
2.1 Reprezentace notovým zápisem	14
2.1.1 Notová osnova	14
2.1.2 Noty a pomlky	15
2.1.3 Notový klíč	15
2.1.4 Předznamenání	16
2.2 Zvuk	16
2.2.1 Tónové třídy	17
3 Zpracování audio signálu	18
3.1 A/D převodník	18
3.2 Formáty audio záznamu	20
4 Základní parametry hudby	22
4.1 Barva zvuku	22
4.1.1 Hudební pohled	22
4.1.2 Technický pohled	22
4.2 Dynamika	23
4.2.1 Hudební pohled	23
4.2.2 Technický pohled	24
4.3 Tempo	26
4.3.1 Hudební pohled	26
4.4 Ladění	28
4.4.1 Hudební pohled	28
4.4.2 Technický pohled	28
5 Synchronizace audio nahrávek	29
5.1 Chroma parametr	29
5.2 Dynamické borcení časové osy	30
5.2.1 Základní princip DTW	30
5.2.2 Deformační cesta	31

6	MIR parametry	33
6.1	Nízkoúrovňové parametry	33
6.1.1	Nízkoúrovňové parametry časové oblasti	33
6.1.2	Nízkoúrovňové parametry frekvenční oblasti	34
6.2	Parametry popisující barvu zvuku	35
6.2.1	Parametry popisující křivku počátků zvukových událostí . . .	35
6.2.2	Parametry popisující rozložení energie ve spektru	36
6.2.3	Drsnost spektra	36
6.2.4	Parametry spektrálního rozložení	37
6.2.5	Melovské keprstrální koeficienty	38
6.3	Parametry popisující dynamiku hudební nahrávky	39
6.3.1	Efektivní hodnota signálu	39
6.3.2	Segmentace signálu	39
6.3.3	Parametr nízké energie	39
6.4	Parametry popisující tempo hudební nahrávky	39
6.4.1	Detekce počátku zvukové události	39
6.4.2	Hustota zvukových událostí	40
6.4.3	Fluktuace tempa	40
6.4.4	Tempo	40
6.4.5	Metrický centroid	40
6.4.6	Metrická síla	40
7	Statistické parametry	41
7.1	Střední hodnota	41
7.2	Směrodatná odchylka	41
7.3	Rozptyl	41
7.4	Kvantily	41
7.4.1	Medián	41
7.4.2	Kvartil	42
7.4.3	Mezikvartilové rozpětí	42
7.4.4	Percentil	42
7.4.5	Mezipercentilové rozpětí	42
8	Databáze interpretací	43
8.1	Rozbor skladby	43
8.1.1	První věta (<i>Allegro con brio</i>)	43
8.2	Databáze	44
8.3	Výběr úseků k analýze	44

9	Vyhodnocovací systém	46
9.1	Předzpracování dat	47
9.2	Výběr stejných úseků	47
9.3	Výpočet parametrů	48
9.4	Výběr parametrů	49
9.4.1	Metoda mRMR	50
9.4.2	Metoda SFFS	50
9.5	Klasifikace	51
9.5.1	Křížová validace	51
9.5.2	Klasifikační stromy	51
9.6	Statistická analýza	51
9.6.1	Přesnost	52
9.6.2	Spearmanův test pořadové korelace	52
9.6.3	Vzájemná informace	52
10	Výsledky	54
10.1	Individuální analýza parametrů	54
10.2	Kombinace parametrů	56
11	Závěr	58
	Literatura	60
	Seznam symbolů, veličin a zkratk	63
	Seznam příloh	64
A	Obsah přiloženého CD	65
B	Databáze nahrávek	66
C	Notový zápis prvního úseku	68
D	Notový zápis druhého úseku	70

SEZNAM OBRÁZKŮ

2.1	Obrázek notové osnovy	14
2.2	Obrázek notových klíčů	16
3.1	Příklad záznamového řetězce	18
4.1	Spektrum tónu C_1 klavíru	23
4.2	Křivky stejné hlasitosti převzato z [13].	26
5.1	Zobrazení příkladů porušení podmínek deformační cesty	32
6.1	Amplitudová obálka audio signálu	33
9.1	Blokové schéma vyhodnocovacího systému	46
9.2	Grafické znázornění vzájemné informace	53

SEZNAM TABULEK

2.1	Tabulka hodnot not a jejich značek	15
2.2	Tabulka hodnot pomlk a jejich značek	15
2.3	Tabulka tónových tříd	17
3.1	Tabulka vzorkovacích kmitočtů	19
3.2	Tabulka dynamických rozsahů pro různé bitové hloubky	20
4.1	Tabulka hudebních dynamik	24
4.2	Tabulka příkladů zvuků a jejich hodnot	25
4.3	Tabulka temp a jejich hodnoty v BPM	27
4.4	Tabulka jazzových tempových změn	27
9.1	Přehled statistických parametrů	49
10.1	Tabulka výsledků individuální analýzy všech parametrů prvního úseku	54
10.2	Tabulka výsledků individuální analýzy všech parametrů druhého úseku	55
10.3	Tabulka parametrů vybraných metodou SFFS pro analyzovaný první úsek	56
10.4	Tabulka parametrů vybraných metodou SFFS pro analyzovaný druhý úsek	56
10.5	Tabulka parametrů vybraných metodou SFFS pro oba analyzované úseky	57
B.1	Tabulka nahrávek v databázi	66
B.1	Tabulka nahrávek v databázi	67

ÚVOD

Diplomová práce se věnuje výzkumu poměrně mladé vědecké oblasti s názvem MIR – *Music information retrieval*. Cílem práce je implementovat vyhodnocovací systém, který za pomoci metod strojového učení a korelační analýzy vybere takové hudební parametry, které mají největší schopnost diferencovat období vzniku jednotlivých interpretací.

Název MIR by se dal volně přeložit do českého jazyka jako získávání informací z hudby. Tento vědecký obor spojuje více vědeckých oborů například: muzikologii, psychologii, zpracování signálů a strojové učení. Cílem MIR je extrahovat z hudebních nahrávek takové parametry, dle kterých lze lépe indexovat hudební soubory, automaticky spravovat databáze hudebních nahrávek nebo zefektivnit nástroje pro vyhledávání hudebních souborů. Tato oblast se také zabývá vývojem algoritmů pro automatické rozpoznání žánrů nebo rozpoznání emocí hudebních nahrávek.

Práce by se dala rozdělit na dvě části. V první části jsou popsány reprezentace jakých může hudba nabývat a jakým způsobem jsou hudební nahrávky získávány. Dále jsou popsány nejpodstatnější parametry, které lze extrahovat z hudebních nahrávek. Tyto parametry jsou rozděleny do skupin podle toho, kterou vlastnost hudebního signálu popisují.

Druhá část této práce je věnována popisu implementovaného vyhodnocovacího systému. Jako vstupní data tohoto vyhodnocovacího systému slouží vytvořená databáze zvukových nahrávek první věty Beethovenovy páté symfonie. Tato databáze obsahuje 56 nahrávek, které byly rozděleny do tří časových kategorií. K analýze byly vybrány dva kontrastní úseky. Z každého úseku je vypočítáno 340 parametrů. Ty jsou následně podrobeny individuální analýze a je vyhodnocena jejich přesnost klasifikace analyzovaných úseků do správných časových kategorií. Následně jsou pomocí metody SFFS vyhledány kombinace parametrů, pro zlepšení přesnosti klasifikace.

1 ZÍSKÁVÁNÍ INFORMACÍ Z HUDEBNÍCH NAHRÁVEK

Získávání informací z hudebních nahrávek nebo-li *Music information retrieval* - MIR je poměrně mladý vědecký obor, který spojuje více vědeckých oborů jako jsou: muzikologie, psychologie, zpracování signálů a strojové učení. Cílem je získávání relevantních dat z hudebních nahrávek pro výzkum a vývoj aplikací. Reálný příklad využití MIR je například doporučení písní, třídění a vyhledávání v online hudebních přehrávačích jako jsou například Spotify nebo Youtube.

1.1 Motivace

Hudba je všudypřítomným tématem v naší společnosti a prakticky každý si užívá ji poslouchat a mnozí ji vytvářet. Obecně řečeno se oblast výzkumu MIR zabývá extrakcí a odvozováním smysluplných dat z audio souborů, indexováním hudby a zlepšení nástrojů pro vyhledávání hudebních souborů. Vzhledem k významu hudby v naší společnosti je až obdivuhodné, že MIR je poměrně mladým vědeckým oborem s původem před méně než dvěma dekádami. Nutno podotknout, že MIR zažívá rostoucí trend jako oblast výzkumu. Jedna z hlavních příčin toho trendu je zlepšení technik komprimace audio souborů v pozdějších 90. letech 20. století, zvětšující se výpočetní výkon počítačů, silné rozšíření mobilních hudebních přehrávačů a vznik streamingových hudebních služeb jako je např. Spotify, Youtube apod. [5, 23]

1.2 Historie MIR

MIR se začal vyvíjet v 90. letech 20. století. Oficiální zrod MIR bychom mohli datovat do roku 2000 na historicky prvním symposiu MIR, které se uskutečnilo v Plymouthu ve státě Massachusetts (USA) pod záštitou mezinárodní organizace ISMIR - The International Society for Music information Retrieval. Ovšem nutno podotknout, že práce které dnes řadíme do oboru MIR se objevovaly již před tímto datem avšak tyto výsledky byly publikovány v různých komunitách a konferencích. Finanční prostředky na toto symposium byly poskytnuty Národní vědeckou nadací (NSF - *National Science Foundation*) prostřednictvím grantového programu zaměřeného na digitální knihovny. Ovšem původní organizátoři Don Byrd, J. Stephen Downie a Tim Crawford byli zaměřeni spíše na symbolickou algebru a digitální knihovny, ale naštěstí symposium přilákalo účastníky, kteří se věnovali práci se zvukem. Interakcí těchto různých oborů vznikl multioborové vědecké odvětví - MIR. [5, 23, 28]

2 REPREZENTACE HUDBY

Hudba může být popsána (reprezentována) mnoha různými způsoby a formáty. V této kapitole budou popsány způsoby jakým může být hudba reprezentována.

2.1 Reprezentace notovým zápisem

Notový zápis poskytuje skrze hudební symboly a značky muzikantovi informaci o tom, jak má být dané dílo interpretováno. Tyto instrukce mohou být znázorněny graficky nebo textově. Původním médiem této notového zápisu je papír, ačkoliv v dnešní době může být notový zápis i podobě digitální. Nutno podotknout, že schopnost správně číst notový zápis není to jediné co musí hudebník zvládnout. Při interpretaci skladby je také důležité hudební cítění muzikanta jímž může utvářet skladbu jako takovou, změnou tempa, dynamikou přednesu nebo artikulací. Nechápejme tedy notový zápis jako přesný postup jak má být skladba interpretována. Chápejme jej spíše jako určitý návod, který nechává prostor pro tvořivost muzikanta, který skladbu interpretuje. [14]

2.1.1 Notová osnova

Základním kamenem každého notového zápisu je tzv. **notová osnova** (viz kapitola 2.1.2). Skládá se z pěti linek. Noty se umísťují buďto na její linky nebo do mezer mezi linkami. Pokud je nota psána nad/pod notovou osnovou užívá se tzv. pomocných linek, které zlepšují přehlednost notového zápisu. Na začátku každé notové osnovy najdeme **notový klíč** (viz kapitola 2.1.3), **předznamení** (viz kapitola 2.1.4) a **metrum**. Pro některé hudební nástroje (např. klavír) potřebujeme více notových osnov. V takovém případě je opatříme tzv. **svorkami**. Na obrázku 2.1 můžeme vidět 5 prázdných taktů notové osnovy pro klavír. V horní osnově můžeme vidět houslový klíč a ve spodním klíči basový. Na konci notové osnovy je finální taktová čára a před prvním taktem je naznačena tónina A dur. Metrum je 4/4. [15]



Obr. 2.1: Obrázek notové osnovy

2.1.2 Noty a pomlky

Nota jako taková se skládá ze tří částí:

- **praporec** - určuje rytmickou hodnotu noty. Může být jednoduchý, dvojitý, trojitý, čtyřnásobný - dle délky noty. V případě, že jsou dvě noty stejné rytmické hodnoty vedle sebe spojují se pomocí **trámců**.
- **nožička** - podstatný je její směr. Od třetí linky, směřují-li noty nahoru, mají noty nožičku směrem dolů. Naopak noty směřující od třetí linky dolů, mají nožičku směrem nahoru. Nožičku píšeme vpravo od hlavičky.
- **hlavička** - má několik základních tvarů. Může být prázdná (u celých a půlových not) nebo plná (v případě ostatních, menších hodnot). Pro bicí nástroje a perkuse se používají speciální tvary hlaviček.

Rozlišujeme několik základních hodnot not, tedy délku jejich trvání - viz tabulka 2.1.

Tab. 2.1: Tabulka hodnot not a jejich značek

Nota	celá	půlová	čtvrtová	osminová	šestnáctinová
Značka					

Pomlky jsou další důležitým prvkem každého notového zápisu. Mají stejné rytmické hodnoty jako noty, ale mají své symboly - viz tabulka 2.2.

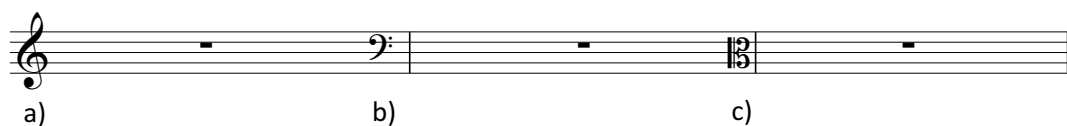
Tab. 2.2: Tabulka hodnot pomlky a jejich značek

Pomlka	celá	půlová	čtvrtová	osminová	šestnáctinová
Značka	-	-	z	7	ž

Můžeme se setkat také se symbolem **tečky** u noty či pomlky. To znamená, že nota či pomlka trvá o polovinu své hodnoty déle. Pro příklad z noty půlové trvajících běžně dvě doby, se stane půlová nota s tečkou a její trvání bude rovno dobám třem. [15]

2.1.3 Notový klíč

Notový klíč je symbol, který určuje pozici určité noty v notové osnově a od něj se odvozuje zápis všech not ostatních. Nejvíce používaný je **klíč houslový** (viz obr. 2.2 a)), také nazývaný **G klíč** a určuje, že nota g^1 je umístěna na druhou linku zespoda. Dále je hojně používaný **klíč basový** (viz obr. 2.2 b)), také zvaný **F klíč**. Ten umísťuje notu *malé f* na čtvrtou linku zespodu. Používaný je také **klíč altový**, někdy nazývaný jako **violový**. Jedná se o **C klíč** (viz obr. 2.2 c)), tedy střed grafického symbolu označuje tón c^1 . [15]



Obr. 2.2: Obrázek notových klíčů

2.1.4 Předznamení

Předznamení značí v jaké tónině je skladba napsána a značí se tzv. **posuvkami**. Nejdříve s tím tedy popíšeme co přesně posuvky značí. Zvyšování/snižování not se značí speciálními znaky - **posuvkami**. O půl tónu směrem nahoru se noty zvyšují pomocí symbolu **křížku** - \sharp a pro snižování noty o půl tónu směrem dolů používáme znak **béčka** - \flat . Takto zvýšený/snížený tón platí až do konce konkrétního taktu. Pokud bychom se chtěli vrátit k tónu původnímu, museli bychom ho opatřit znakem **odrážky** - \natural . Odrážka platí i pro zrušení předznamení. Odrážka umístěná před notou ruší předznamení pod dobu jednoho taktu.

Předznamení se píše na začátek notové osnovy. Pro příklad mějme notový zápis s předznamením tří křížků, což znamená že skladba je v tónině A dur. Tato tónina obsahuje křížky $f\sharp$, $c\sharp$, $g\sharp$. Výchozí nastavení takto udané tóniny značí, že pokud v notaci uvidíme tóny f , c , g hrajeme je automaticky s příslušnými posuvkami. Pokud na začátku notové osnovy není uvedeno předznamení žádné, je daná skladba tónině C dur. Ta totiž neobsahuje žádné tóny s posuvkami. [15]

2.2 Zvuk

Hraní na jakýkoliv akustický hudební nástroj vytváří zvuk. Jedná se o mechanické vlnění v libovolném pružném hmotném prostředí. Toto prostředí může být skupenství plynného, kapalného nebo pevného. V našem případě uvažujeme prostředí plynné tedy vzduch. Toto mechanické vlnění, vytvořené libovolným zdrojem zvuku např. hudebním nástrojem, má za následek lokální zhušťování a zředování molekul vzduchu - vytvoří se tedy zvuková vlna. V případě, že tato vlna dorazí k akustickému přijímači např. lidskému uchu, vyvolá tato zvuková vlna změnu rovnovážné polohy ušního bubínku. Tuto změnu potom vnímáme jako zvuková vjem. Rychlost zvuku je za normální teploty vzduchu $t = 20^\circ\text{C}$ a normálním atmosférickým tlakem $p_0 = 101,3\text{ kPa}$ zhruba $c_0 = 343\text{ m/s}$. Udává se, že lidský sluch je schopen vnímat zvuky o frekvencích v rozmezí $20\text{ Hz} - 20\text{ kHz}$. Toto rozmezí je ale velice individuální a pravý rozsah závisí na věku jedince. Podstatnou roli hrají také jakým zvukům či hlukům je jedinec vystavován. Nejcitlivější je lidský sluch na zvuk v oblasti $f = 2 - 4\text{ kHz}$ (viz obr. 4.2). [14, 24]

Zvuky můžeme rozlišit na dvě kategorie:

- **Tóny** vznikají pravidelným, v čase přibližně periodickým, kmitáním. Při jejich poslechu vzniká zvukový vjem u něhož můžeme určit určitou výšku. Zdrojem tónů mohou být např. lidské hlasivky (zpěv) nebo různé hudební nástroje. Rozlišujeme 12 tónových tříd viz kapitola 2.2.1.
- **Hluky** jsou nepravidelné zvukové vlnění. Tyto hluky mohou vznikat např. srážkou dvou těles apod. Za jistý typ hluku můžeme považovat i zvuk bicích nástrojů.

Každý zvuk nebo zvukové vlnění můžeme popsat řadou parametrů. My si uvedeme tři z nich, abychom získali představu jakými parametry zvuk popisujeme. Budou to:

- **Frekvence** - udává počet opakování zvukové vlny za daný časový úsek. Explicitně je definována výrazem $f = \frac{1}{T}$, kde T je časová perioda. U zvuku spíše používáme definici následující:

$$f = \frac{c_0}{\lambda} \quad [\text{Hz}], \quad (2.1)$$

kde c_0 je rychlost zvuku a λ je vlnová délka.

- **Hlasitost** - je subjektivní veličina, která je závislá na velikosti akustického tlaku p . Blíže je tato veličina popsána v kapitole 4.2.2.
- **Barva** - pomocí ní jsme schopni rozlišit zvuk např. klarinetu od flétny apod. Tato skutečnost je způsobena přítomností harmonických složek ve spektru zvuku. Blíže bude popsána v kapitole 4.1.2.

2.2.1 Tónové třídy

Mezi jednotlivými tóny můžeme rozlišovat tzv. intervaly. Asi nejjednodušším příkladem intervalu mezi dvěma tóny je oktáva, jejichž frekvenční poměr je 1:2. Oktáva je rozdělena na 12 stejně velkých částí viz tabulka 2.3. [14, 20]

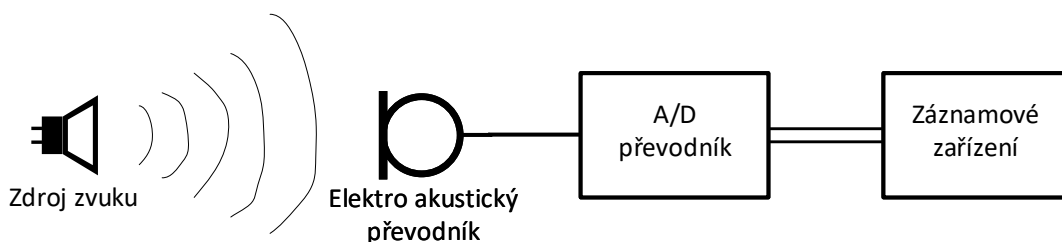
Tab. 2.3: Tabulka tónových tříd

Tón. tř.	0	1	2	3	4	5	6	7	8	9	10	11
Tón	C	C \sharp /D \flat	D	D \sharp /E \flat	E	F	F \sharp /G \flat	G	G \sharp /A \flat	A	A \sharp /H \flat	H

3 ZPRACOVÁNÍ AUDIO SIGNÁLU

V této kapitole si popíšeme jakým způsobem je zvuk zpracováván a jaká média se pro uchování zvukové informace běžně používají. Nejprve si ale musíme říci, jaké kroky je potřeba provést, než jsme schopni zvukový záznam zpracovávat a následně ukládat na libovolné médium. Na obrázku 3.1 můžeme vidět jednoduchý příklad záznamového řetězce. Tento řetězec se skládá ze:

- **zdroje zvuku** generuje zvukové vlnění, tedy zvuk, který chceme zaznamenat. Toto vlnění se šíří prostředím (vzduchem) až dorazí k elektroakustickému převodníku. Pod pojmem zdroj zvuku si můžeme představit např. hudební nástroj, zpěvák, reproduktor apod.
- **elektroakustický převodník** slouží k převodu zvukového vlnění na elektrické napětí, které dále zpracováváme. Příklad takového převodníku je mikrofón.
- **A/D převodník** (*Analogově digitální převodník*) slouží k převodu elektrického napětí na digitální data, v jistém formátu. V tomto případě je typický A/D převodník zvuková karta (více v kapitole 3.1).
- **záznamové zařízení** - zde ukládáme pořízený zvukový záznam v určitém formátu dat (více v kapitole 3.2). Může to být např. počítač.



Obr. 3.1: Příklad záznamového řetězce

3.1 A/D převodník

Jak již bylo naznačeno výše, úkolem tohoto zařízení je převod elektrického spojitého signálu na digitální data. Tento proces se obecně skládá ze 4 operací:

- **Úprava vstupního signálu** - aby nedošlo ke zkreslení zpracovávaného signálu, ale zároveň byla využita maximální rozlišovací schopnost A/D převodníku, je potřeba velikost vstupního napětí patřičně zesílit/zeslabit. Tato operace je provedena pomocí napětového zesilovače. Jeho činnost je buď to řízena uživatelem (ve většině případů se jeho zesílení volí potenciometrem s popisem

Gain) nebo je jeho činnost řízena automaticky patřičným obvodem (mikropočítačem). Dále je potřeba splnit podmínku *Shannonova-Nyquistova teorému*. Ten je definován jako:

$$f_{VZ} > 2 \cdot f_{MAX}, \quad (3.1)$$

kde f_{VZ} je vzorkovací kmitočet a f_{MAX} je nejvyšší harmonická složka vstupního signálu. Tato poučka nám tedy říká, že vzorkovací kmitočet musí být alespoň dvojnásobný než kmitočet nejvyšší harmonické složky vstupního signálu. Při nedodržení této poučky by docházelo při rekonstrukci signálu ke značnému zkreslení tvaru signálu. Tato podmínka je splněna zařazením filtru typu dolní propust do cesty vstupního signálu. Její mezní kmitočet a strmost je zvolena v závislosti na hodnotě vzorkovací kmitočtu. [25]

- **Vzorkování vstupního signálu** - při převodu analogového signálu na digitální je nutné provést vzorkování zpracovávaného signálu, jelikož vstupní analogový signál může nabývat nekonečně mnoha hodnot. Tato operace tedy spočívá ve „vyseknutí“ vzorku vstupního signálu. Tyto vzorky jsou odebírány s jistou frekvencí f_{VZ} , kterou určuje oscilátor. Pro audio aplikace se volí nejčastěji tyto vzorkovací kmitočty:

Tab. 3.1: Tabulka vzorkovacích kmitočtů

Vzorkovací frekvence f_{VZ} [kHz]	44,1	48,0	88,2	96,0	192,0
Vzorkovací perioda T_{VZ} [ms]	22,68	20,83	11,34	10,42	5,21
Počet vzorků za 1 sekundu	44100	48000	88200	96000	192000

Nutno podotknout, že po vzorkování zpracovávaného signálu došlo k jeho diskretizaci v časové oblasti. Hodnota vzorků je ovšem stále spojitá a je potřeba se s tímto faktem vypořádat určitý způsobem. K tomu slouží proces kvantizace.

- **Kvantování navzorkovaného signálu** má za úkol zaokrouhlit spojitou hodnotu získaných vzorků na definované diskrétní hodnoty tzv (*kvantovací hladiny*) a jejich počet je definován počtem bitů následného kódování n (tzv. *bitové rozlišení*). Můžeme definovat tzv. kvantovací krok Q :

$$Q = \frac{1}{(2^n - 1)}, \quad (3.2)$$

kde n je počet bitů převodníku. Můžeme také definovat kvantovací chybu, tedy rozdíl původního nakvantovaného signálu. Jedná se o chybu způsobenou diskrétními úrovněmi výstupního signálu kvantovače. Maximální odchylka převodníku je dána polovinou hodnoty nejmenšího kvantovacího kroku. Bývá definován také kvantovací šum, který představuje odchylku nakvantovaného

signálu oproti původnímu. Pro harmonické signály je kvantovací šum definován rovnicí:

$$SNR = 6,02n + 1,76 \quad [\text{dB}]. \quad (3.3)$$

Někdy se používá zjednodušený odhad kvantovacího šumu, který je definován jako:

$$SNR = 20 \log_{10} \frac{1}{Q} \quad [\text{dB}]. \quad (3.4)$$

Pro audio aplikace se užívá bitového rozlišení 16, 20 a 24 bitů. Bitové rozlišení nám také definuje dynamický rozsah převodníku a je dán vztahem:

$$DR = 20 \log_{10} \left(\frac{2^Q}{1} \right) = 6,02 \cdot Q \quad [\text{dB}]. \quad (3.5)$$

V tabulce 3.2 jsou uvedeny, výpočty dynamických rozsahů pro používané bitové hloubky v audio aplikacích. Standardem je formát CD, který pracuje se vzorkovací frekvencí $f_{VZ} = 44,1$ [kHz] a s bitovou hloubkou $Q = 16$ bitů. [25]

Tab. 3.2: Tabulka dynamických rozsahů pro různé bitové hloubky

Q [bit]	16	20	24
DR [dB]	96	120	145

- **Kódování** je proces, při kterém převádíme vzorky s diskretním časem a s diskretní úrovní na binární číslo. V audio aplikacích, v profesionální sféře, se využívá výhradně kódování PCM - *Pulse Code Modulation*. Výhoda toho kódování je v tom, že nezpůsobuje ztrátu informace nebo omezení šířky pásma rychlých přechodů. Také umožňuje přesnost nástřihu na jeden vzorek bez nutnosti překódování. V digitálních audio systémech se využívá pro výpočet úrovně signálu v dB relativní jednotka úrovně dBFS (dB to relative Full Scale). Nominální úroveň je zde definována jako 0 dBFS, která odpovídá maximální hodnotě digitálního signálu, který může převodník zpracovat. Systémy které pracují s čísly s plovoucí nebo fixní desetinnou čárkou ve zlomkovém formátu odpovídá 0 dBFS číslo 1. Tyto systémy např. jsou: osobní počítače, signálové procesory apod. [25]

3.2 Formáty audio záznamu

Zvuk ukládáme do záznamového zařízení v různých formátech. Často jsou tyto formáty komprimovány, pro omezení jejich velikosti. Podle typu komprese rozlišujeme tři základní druhy audio formátů:

- **Audio formáty bez komprese** - tyto formáty nepoužívají kompresi a data jsou kódovány v PCM. Obvykle je takovéto audio data ukládána do souborů s koncovkou `.wav` ve Windows nebo `.aiff` pro MAC OS. Oba tyto typy souborů obsahují v datovém řetězci hlavičku ve které jsou uloženy informace o délce záznamu, počtu kanálů, použité vzorkovací frekvenci a bitové hloubce.
- **Audio formáty s bezztrátovou kompresí** - tyto audio formáty snižují náročnost na velikost audio souborů, za použití bezztrátové komprese. Typickým typem jsou soubory s koncovkou `.flac` nebo `.m4a`. Tento typ souborů používá lineární predikci pro konverzi zvukových vzorků do série malých čísel, která jsou efektivně kódována pomocí *Golom-Riceova kódování*.
- **Audio formáty se ztrátovou kompresí** - nejtypičtější představitelem těchto formátů jsou soubory s koncovkou `.mp3`. Tento formát využívá nedokonalosti lidského sluchu, princip tzv. maskování. Odstraňuje tedy informace, které člověk neslyší nebo je není schopen postřehnout. Tím se výrazně snižují nároky na velikost souboru. Pro příklad: audio soubor CD kvality je možné ve formátu `.mp3` uložit s velikostí jedné desetiny oproti velikosti původní.

4 ZÁKLADNÍ PARAMETRY HUDBY

V této kapitole budou popsány základní parametry hudby, které mají zásadní vliv na hudební projev skladby. Každý parametr bude popsán ze dvou pohledů. Z pohledu hudebního, kde bude rozebráno jakým způsobem je vybraný parametr popisován hudebně a jak je zapsán v notovém zápise a z pohledu technického, kde bude popsáno jaké technické veličiny vybranému parametru přisuzujeme a jak jsou definovány.

4.1 Barva zvuku

Barva zvuku umožňuje posluchači rozeznat zvuk jednotlivých nástrojů od sebe i v případě, že nástroje hrají ten samý tón se stejnou hlasitostí. Stejně jako hlasitost je barva zvuku subjektivní parametr. Neexistují tedy sjednocené názvosloví pro různé barvy zvuků. Typické slovní označení pro barvu zvuku je např. teplá, ostrá, plná apod. [14, 27]

4.1.1 Hudební pohled

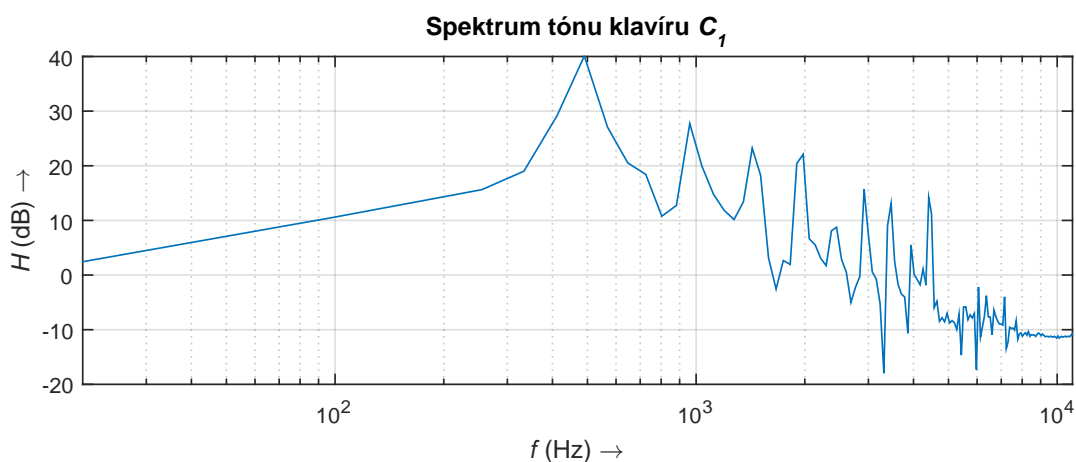
Z hudebního hlediska barva zvuku hudebních nástrojů hraje zásadní roli v instrumentaci. Pomocí míchání zvukových barev jednotlivých nástrojů vznikají zvukové mixtury. Volbou vhodných hudebních nástrojů v instrumentaci můžeme ovlivnit emoční náladu celé skladby nebo její části. Nutno podotknout, že vliv na emoční náladu skladby nebo její části má také zvolená tónina. [15]

4.1.2 Technický pohled

Barvu zvuku určuje tvar vlny jakou daný zvuk generuje. Tuto vlnu můžeme analyzovat z pohledu frekvenčního nebo časového (tedy její obálku). Oba tyto pohledy jsou vzájemně provázány. To znamená, že pokud zanalyzujeme jistou zvukovou vlnu frekvenčně, jsme schopni odhadnout jaký obálku zvuk má. [27]

Frekvenční hledisko

Bohatost spektra zvuku nebo tónu hudebního nástroje závisí na počtu harmonických složek, které hudební nástroj generuje. Harmonickou složku s nejnižší kmitočtem nazýváme základní harmonickou složkou a její frekvence definuje tón zvuku. Ostatní harmonické složky tvoří barvu zvuku a jejich frekvence jsou celočíselné násobky základní harmonické složky. Nástrojem pro frekvenční analýzu obecného signálu je FFT - (*Fast Fourier Transform* - rychlá Fourierova transformace). Následné grafické vyjádření může být provedeno např. pomocí vyobrazení spektra zvuku v závislosti na frekvenci (viz obr. 4.1). [27, 21]



Obr. 4.1: Spektrum tónu C_1 klavíru

Časové hledisko

Barva zvuku je také velmi ovlivněna obálkou vlny, kterou daný hudební nástroj generuje. Obálku popisujeme pomocí tzv. *ADSR* parametrů: **rychlost náběhu** (*attack time*) - jedná se rychlost s jakou daný zvuk dosáhne své maximální hodnoty od počátku, **doba přechodu** (*decay*) - čas za který přejde zvuk z maximální hodnoty do ustáleného stavu, **doba trvání ustáleného zvuku** (*sustain*), **rychlost doběhu** (*release time*) - doba za kterou zvuk zanikne. Tyto čtyři fáze zvuku jsou klíčové pro rozlišení barev jednotlivých nástrojů. Pokud bych např. u stejného tónu klavíru a trumpetu odstranili fázi rychlosti náběhu tónu, rozlišení jednotlivých barev tónů by se razantně zhoršilo. U některých hudebních nástrojů fáze ustálení zvuku úplně chybí. Typicky se toto děje u bicích nástrojů, s minimálním tónovým obsahem (malý buben). Tento jev je způsoben úderem paličky na blánu bubnu. Výsledkem je zvuk se strmým nárůstem do maximální hodnoty a s krátkým časem trvání. Tyto druhy zvuků nazýváme tranzienty. Výsledkem strmého nárůstu je velký počet harmonických složek ve spektru zvuku. [27]

4.2 Dynamika

4.2.1 Hudební pohled

Dynamika přednesu je v hudebním zápise značena tzv. dynamickými značkami. Běžně je používáno 7 úrovní dynamiky viz tabulka 4.1. U těchto značek dynamiky mohou být také naznačeny akcenty, značeny přidáním písmena *s* nebo vložení značky dynamiky mezi dvojici písmen *s* a *z*. Pro příklad *sf* (*sforzando*) nebo *sfz* (*sforzato*). Je zřejmé, že tyto instrukce o dynamice jsou velmi neurčité a jejich vjem závisí čistě na subjektivním citění hráče. V důsledku to znamená, že např. *pp* hráče

A nemusí být stejné intenzity jako hráče B.

Dále může být v notovém zápise uvedena značka plynulé změny dynamiky. Graficky se tato změna značí tzv. *vidlicemi* $<>$ pod či nad takty, ve kterých se tato změna dynamiky má odehrát. Na začátku a konci *vidlice* bývá ve většině případů uvedeno z jaké úrovně dynamiky do jaké se tato změna má odehrávat. Technicky řečeno na začátku je uveden výchozí bod a na konci bod koncový. [15]

Tab. 4.1: Tabulka hudebních dynamik

Zkratka		význam
<i>ppp</i>	<i>piano pianissimo</i>	co nejslaběji
<i>pp</i>	<i>pianissimo</i>	velmi slabě
<i>p</i>	<i>piano</i>	slabě
<i>mf</i>	<i>mezzoforte</i>	středně silně
<i>f</i>	<i>forte</i>	velmi silně
<i>ff</i>	<i>fortissimo</i>	středně silně
<i>fff</i>	<i>forte fortissimo</i>	co nejsilněji
<i>sf/sfz</i>	<i>sforzando/sforzato</i>	se silným důrazem
<i>cresc.</i> ($<$)	<i>crescendo</i>	zesilovat
<i>decresc.</i> ($>$)	<i>decrescendo</i>	zeslabovat
<i>dim.</i> ($>$)	<i>diminuendo</i>	zeslabovat

4.2.2 Technický pohled

V předešlé kapitole 4.2.1 bylo popsáno jak je dynamika chápána z hlediska hudebního. Nyní si rozebereme jakou veličinou definujeme dynamiku z hlediska technického. Hlasitostí zvuku se rozumí „vlastnost sluchového vnímání, podle které lze uspořádat zvuky na stupnici v rozsahu od tichých po hlasité“, dle normy[4]. Subjektivní hlasitost zvuku souvisí s fyzikální veličinou zvanou intenzita zvuku značena I , která je definována pro složené zvuky jako

$$I = \frac{1}{T_{max}} \int_0^{T_{max}} p v dt \quad [W/m^2], \quad (4.1)$$

kde $T_{max} \gg 1/f_n$ a f_n je kmitočet nejnižší spektrální složky složeného zvuku, p je akustický tlak zvuku a v je akustická rychlost. Rovnice 4.1 je ovšem základní definice. Pro posouzení hlasitosti zvuku se využívá tzv. hladina intenzity zvuku. Tato veličina je poměrová, jejíž referenční hodnotou je I_0 je referenční hodnota $10^{-12} W/m^2$. Pro velký rozsah hodnot je veličina logaritmována. Zavádíme tedy jednotku

decibel ([dB]). Hladina intenzity zvuku je definována

$$L_I = 10 \log_{10} \frac{I}{I_0} \quad [\text{dB}]. \quad (4.2)$$

Z této definice je patrné, že dvojnásobek hladiny intenzity zvuku vede ke zvýšení výsledku o 3 dB:

$$\text{dB}(2 \cdot I) = 10 \cdot \log_{10}(2) + [\text{dB}](I) \approx 3 + \text{dB}(I). \quad (4.3)$$

V tabulce 4.2 můžeme vidět příklady zvuků a jejich hodnoty intenzity a hladiny intenzity zvuku. Povšimněme si, že hodnota intenzity zvuku se od prahu slyšení po práh bolesti změnila o 12 řádů. Zavedení jednotky *decibel* ([dB]) je tedy velmi výhodné vzhledem v jakém rozmezí se hodnota intenzity zvuku mění.

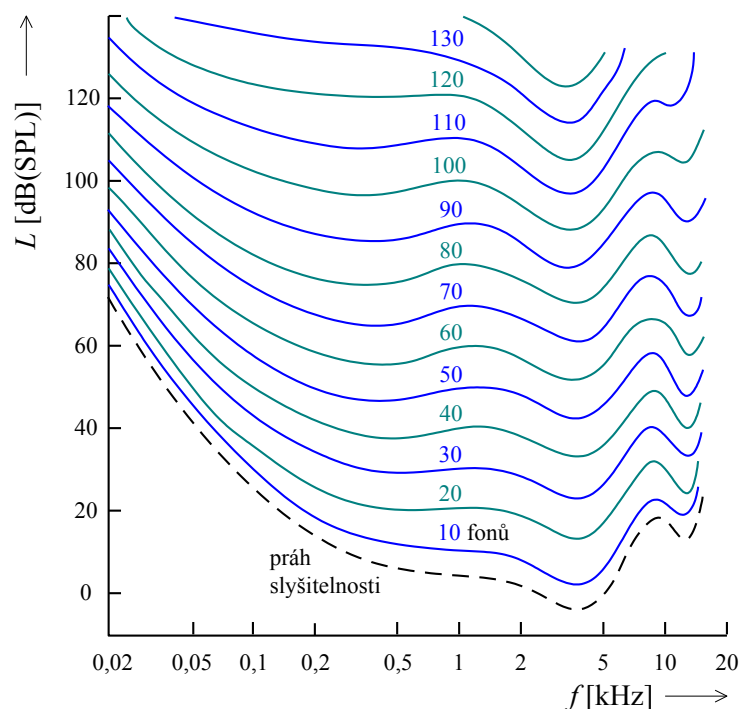
Tab. 4.2: Tabulka příkladů zvuků a jejich hodnot

Zdroj	Intezita [W/m^2]	Hladina intenzity [dB]
Práh slyšení	10^{-12}	0
Šepot	10^{-10}	20
<i>Pianissimo</i>	10^{-8}	40
Běžná konverzace	10^{-6}	60
<i>Fortissimo</i>	10^{-2}	100
Práh bolesti	10	130
Vzlet tryskové letadla	10^2	140

Nutno však podotknout, že tato logaritmická závislost je závislá také na kmitočtu. Subjektivním srovnáváním hlasitostí tónů různých kmitočtů s tónem referenčním o frekvenci $f = 1000$ Hz, byly získány křivky tzv. *křivky stejné hlasitosti* (Fletcherovy-Munsonovy křivky) - viz obr4.2. Tyto křivky udávají jaký akustický tlak způsobí na různých kmitočtech stejně hlasitý vjem jako referenční tón. Jednotkou hladiny hlasitosti je fon [Ph]. Zajímavý a podstatný je fakt, že lidské ucho je nejcitlivější na kmitočty v rozmezí $f = 2-4$ [kHz]. Nutno také podotknout, že frekvenční rozsah lidského ucha je běžně udáván v rozmezí kmitočtů 20 Hz – 20 kHz. Ten se, ale s věkem razantně mění a také záleží jakým a jak hlasitým zvukovým vjemům je lidské ucho vystavováno. [14, 24]

Na dynamiku zvuku je možné také pohlížet z hlediska jiného - signálového. Z tohoto pohledu je dynamický rozsah analogového zařízení definován jako rozdíl mezi maximální RMS hodnotou nezkresleného harmonického průběhu (popřípadě hodnotou RMS s definovaným harmonickým zkreslením) a hladinou vlastního šumu. V takovémto případě udáváme dynamický rozsah opět v decibelech.

Můžeme se také setkat s pojmem *dynamický rozsah zvukového programu*. Za zvukový program je označován ucelený, časově definovaný úsek zvukového signálu.



Obr. 4.2: Křivky stejné hlasitosti převzato z [13].

Dynamický rozsah zvukového programu je poté definován jako rozsah maximálních a minimálních efektivních hodnot, které se zvukovém programu v průběhu času objeví. Můžeme také definovat *nominální úroveň* zvukového programu. Jedná se o dlouhodobě měřenou efektivní hodnotu zvukového programu vyjádřenou v dB. [25]

4.3 Tempo

Jak rychle nebo pomalu má být skladba hrána nám udává tzv. *tempo*, které bývá explicitně uvedeno na začátku skladby nebo na začátku jednotlivých částí. Má také přímý dopad na délku zvukové nahrávky jako takové. Jednoduše řečeno, čím pomalejší tempo hráč zvolí, tím větší čas bude zvuková nahrávka mít.

4.3.1 Hudební pohled

Úzce spjatý s tempem v hudební rovině jsou pojmy **takt**, **metrum** a **taktová čára**. Nejprve si popíšeme tedy je. Hudební zápis je rozdělen do rovnoměrných částí nazývaných **takt**, jím rozumíme časový úsek ohraničený **taktovou čarou**. Způsob počítání dob je určen tzv. **metrem**, které bývá označeno zlomkem např. 4/4. Čítec udává informaci o počtu dob a jmenovatel udává rytmickou hodnotu jednotlivých dob. Doby v taktu dělíme na tzv. **těžké** a **lehké**. Pro příklad, mějme 2/4 takt. První doba je **těžká** a druhá **lehká**.

Označení tempa skladby bývá většinou uvedeno na začátku skladby přednesovým značením italským (viz tabulka 4.3) nebo anglicky (např. *fast*, *medium* apod.) Nepřesnějším označením tempa je však metronomické značení BMP (*Beats per minute* - počet úderů na minutu). Prostor mezi dvěma údery je rytmickou hodnotou noty. Po příklad označení ♩ = 60 znamená, čtvrtová nota bude mít trvání jednu vteřinu. Můžeme se také ještě setkat s označeními zrychlování (***accelerando***) nebo zpomalování (***ritardando***). Značka ***rubato*** značí, že vybraná část skladby má být hrána volněji. Pokud chceme po takové části nastolit tempo, které bylo uvedeno na začátku skladby píšeme značku ***a tempo***.

Tab. 4.3: Tabulka temp a jejich hodnoty v BPM

Značení	Význam	BPM
Pomalé tempo		
<i>Grave</i>	těžce	40
<i>Largo</i>	široce	44
<i>Lento</i>	rozvláčně	50
<i>Adagio</i>	pomalů, zvolna	54
Středně rychlé tempo		
<i>Andante</i>	volně, krokem	63
<i>Andantino</i>	volně, rychleji, než andante	69
<i>Moderato</i>	mírně	88
<i>Allegretto</i>	vesele, mírněji než allegro	106
Rychlé tempo		
<i>Allegro</i>	vesele, rychle	132
<i>Allegro assai</i>	značně rychle	144
<i>Vivace</i>	živě	160
<i>Presto</i>	velmi rychle	184
<i>Prestissimo</i>	co nejrychleji	208

V jazzové hudbě se také můžeme setkat se speciálními tempovými změnami, kdy se určitá část skladby hraje např. v tempu polovičním (*half-time*) nebo v tempu dvojnásobném (*double time*). V takovém případě se, ale nemění počet dob, ale jen hustota rytmu viz tabulka 4.4. [15]

Tab. 4.4: Tabulka jazzových tempových změn

<i>half-time</i>	1				2				
původní tempo	1		2		3		4		
<i>double time</i>	1	2	3	4	1	2	3	4	

4.4 Ladění

4.4.1 Hudební pohled

Laděním určujeme přesné vzdálenosti jednotlivých tónů v hudební stupnici resp. ladění definuje frekvence jednotlivých tónů a jejich poměry mezi nimi. Tyto poměry mezi jednotlivými tóny nazýváme intervaly. Základním intervalem je **oktáva**. Jedná se o vzdálenost mezi základní a první harmonickou složkou a poměr frekvencí je 1:2. Vzdálenost mezi druhou a třetí harmonickou složkou je **kvinta**, s poměrem frekvencí 3:2. Dále mezi třetí a čtvrtou harm. složkou je **kvarta** s poměrem 4:3 atd. Pokud poskládáme tóny za sebe dle jistých pravidel v rozsahu jedné oktávy vznikne hudební stupnice. Nemůžeme ji ovšem vytvořit tak, aby všechny všechny intervaly mezi jednotlivými tóny byly „čisté“ (konsonantní). [15, 21]

Existují dva druhy ladění:

- **přirozená ladění** - u tohoto druhu ladění se používají pouze tóny, jejichž frekvence jsou ve vzájemných poměrech vyjádřitelných celými čísly. Toto ladění má ovšem velké nedostatky. Frekvence jednotlivých tónů se v různých stupnicích liší. Toto je způsobeno tím, že dvanáctý půltón nebude nesouhlasit s čistou oktávou.
- **temperovaná ladění** - toto ladění využívá záměrného mírného rozladění některých čistých intervalů tak, aby se docílilo přesnějšího naladění intervalů jiných. Existují dva druhy temperovaného ladění: nerovnoměrné a rovnoměrné. V současnosti je nejpoužívanější ladění rovnoměrně temperované, kde všechny intervaly jsou stejně rozladěny, ale žádný interval kromě oktáv není „čistý“. Poměr frekvencí jednotlivých půltónů je u tohoto ladění zvolen $\sqrt[12]{2}$.

4.4.2 Technický pohled

Pro měření velikosti intervalů se využívá bezrozměrná jednotka *cent*. Její definice vychází z rovnoměrně temperovaného ladění. Oktáva je tedy rozdělena na 12 stejně velkých půltónů. Jeden *cent* je poté definován jako jedna setina temperovaného půltónu. Jeden cent je tedy definován $1\text{ c} = \sqrt[100]{\sqrt[12]{2}} = \sqrt[1200]{2}$. [14, 21]

5 SYNCHRONIZACE AUDIO NAHRÁVEK

Hudba může být popsána a reprezentována mnoha různými způsoby např. jako notový zápis či audio nahrávka. Pro každou tuto reprezentaci může existovat nespočet různých verzí provedení odpovídající stejnému hudebnímu dílu (skladbě). Tyto různé verze stejného hudebního díla či skladby nazýváme interpretace. Jednotlivé interpretace se mohou lišit jednak tempově, ale také i instrumentálním provedením. Pro příklad Beethovenova pátá symfonie, známá jako „osudová“, je originálně napsána pro symfonický orchestr, ale existují i interpretace např. pro sólový klavír apod. Nemluvě o tom, že každý dirigent symfonického orchestru, ačkoliv je v notovém zápise explicitně uvedeno jakým způsobem má být skladba interpretována, ji diriguje dle svého hudebního citu. Tento fakt má za následek odchylky v interpretacích stejného hudebního díla. Aby bylo možné objektivně porovnat dva vybrané úseky skladby různých interpretací je zapotřebí tyto nuance odstranit a právě toto je cílem synchronizace hudebních nahrávek.

Typicky se synchronizace audio nahrávek děje ve dvou základních krocích. Krok první, obě porovnávané nahrávky jsou převedeny do tvaru, který je výhodný pro další zpracování. V našem případě se využívá převodu vybraných úseků na tzv. chromagramy více v kapitole 5.1. V kroku druhém dochází již k synchronizaci jako takové, za použití nelineární metody - dynamického borcení časové osy (DTW), která bude blíže popsána v kapitole 5.2.

5.1 Chroma parametr

Jedna z vlastností lidského sluchu je ta, že dva tóny, které mají podobnou harmonickou funkci a jejich rozestup je minimálně oktáva, vnímá jako dva tóny s podobnou zvukovou barvou. V tomto smyslu vnímá lidský sluch výšku tónu periodicky. Díky tomuto faktu může být výška tónu rozdělena na dvě součásti: tónová výška a tónová třída. Tónovou výškou popisujeme v jaké oktávě se daný tón nachází a tónová třída udává do jaké tónové třídy daný tón patří. *Chroma parametr* zapouzdřuje tyto dvě dílčí součásti do jedné reprezentace. Jedná se tedy o vektor o 12 dimenzích, kde jedna dimenze reprezentuje jeden tón tónových tříd. Cílem *chroma parametru* je seskupení všech spektrálních informací vztahujících se ke konkrétní tónové třídě pod jedním koeficientem. V praxi je *chroma parametr* používán ve formě tzv. *chromagramu*, který je vytvořen aplikováním FFT. Analogickou reprezentací *chromagramu* je *spektrogram*, ovšem s tím rozdílem, že osa y (frekvenční) je u *chromagramu* reprezentována tónovými třídami. *Chromagramy* se staly základním kamenem při úkolech synchronizace hudební nahrávek nebo při rozpoznávání akordů. [14, 29]

5.2 Dynamické borcení časové osy

Na začátku kapitoly 5 jsme si nastínili jak odlišné mohou interpretace stejného hudebního díla být. Také jsme si ukázali v podkapitole 5.1 do jaké podoby porovnávání hudební nahrávky převádíme s ohledem na to, aby další operace („zarovnání“) byla efektivní. Nyní si podrobně rozebereme, jak ono zarovnání, tedy synchronizace hudebních nahrávek, se děje. Využívá se k tomu nelineární operace tzv. dynamické borcení časové osy - DTW (*Dynamic time wrapping*). Cílem metody DTW je tedy „zarovnání“ dvou jakýchkoliv časově závislých posloupností dat za určitých předpisů. Původně byla tato metoda DTW využívána ve zpracování řeči v oblasti automatické klasifikace slov. Nyní je využívána také při tzv. dolování dat (*data mining*), v bioinformatice a také v MIR.

V našem případě tedy využíváme metodu DTW pro zarovnání dvou hudebních nahrávek, které chceme vzájemně porovnat. Nastíňme si následující scénář. Mějme dvě chroma posloupnosti, dvou odlišných interpretací stejného hudebního díla. První posloupnost je definována jako $X = (x_1, x_2, x_3, \dots, x_N)$, mající délku $N \in \mathbb{N}$ a druhá $Y = (y_1, y_2, y_3, \dots, y_M)$, mající délku $M \in \mathbb{N}$, kde prvky x_N a y_N jsou vektory dvou chroma posloupností. Pro příklad, posloupnost X je chroma vektor interpretace Beethovenové páté symfonie, hrající symfonický orchestr a posloupnost Y je chroma vektorem toho samého hudebního díla, ale ve verzi sóla pro klavír. Jelikož dirigent symfonického orchestru udal celkové tempo skladby větší než hráč piána, nechť má posloupnost X délku $N = 12$ a posloupnost Y délku $M = 15$. Abychom mohli tedy tyto dvě posloupnosti porovnat musíme aplikovat nelineární operaci tak, aby posloupnost X měla stejnou délku jako posloupnost Y , tedy $N = M$. Intuitivně by toho šlo dosáhnout vynecháním určitých prvků z posloupnosti Y nebo použitím určitých prvků posloupnosti X více než jednou. [14]

5.2.1 Základní princip DTW

Jak bylo řečeno výše, hlavním cílem DTW je porovnání dvou posloupností například $X = (x_1, x_2, x_3, \dots, x_N)$, mající délku $N \in \mathbb{N}$ a $Y = (y_1, y_2, y_3, \dots, y_M)$, mající délku $M \in \mathbb{N}$. Pomineme-li scénář porovnávání dvou audio nahrávek, tyto posloupnosti mohou reprezentovat diskrétní signál, sekvenci znaků, jednoduše jakoukoliv časově závislou posloupnost.

Definujme si prostor F a předpokládejme, že $x_n, y_m \in F$ pro $n \in [1 : N]$ a $m \in [1 : M]$. Pro porovnání dvou různých funkcí $x, y \in F$ potřebujeme funkci tzv. *local cost measure*, která je definována jako

$$c : F \times F \rightarrow \mathbb{R}. \quad (5.1)$$

Lze tvrdit, že $c(x, y)$ je malé - má malou cenu (*cost*), pokud x a y jsou si velmi podobné. V opačném případě $c(x, y)$ je velké nebo-li má velkou cenu. Vyhodnocením *local cost measure* pro každý prvek posloupností X a Y je získána tzv. matice ceny (*cost matrix*) $\mathbf{C} \in \mathbb{R}^{N \times M}$ definována

$$\mathbf{C}(n, m) := c(x_n, y_m), \quad (5.2)$$

kde platí $n \in [1 : N]$ a $m \in [1 : M]$. Pro další zjednodušení, pár souřadnic (n, m) matice ceny \mathbf{C} definujeme jako buňku matice \mathbf{C} .

V případě synchronizace audio nahrávek bude prostoru F náležet vektor chroma posloupností o dvanácti dimenzích, tedy $F = \mathbb{R}^{12}$. Existuje mnoho způsobů jak určit vzdálenost mezi dvěma prvky $x, y \in F$. Asi nejznámější je využití Euklidovské vzdálenostní funkce definována jako

$$c(x, y) = \sqrt{\sum_{l=1}^K (x_l - y_l)^2}, \quad (5.3)$$

kde K reprezentuje počet dimenzí, v našem případě tedy $K = 12$.

Další využívanou vzdálenostní funkcí je Kosinova vzdálenostní funkce, která je definována:

$$c(x, y) = 1 - \frac{\langle x | y \rangle}{\|x\| \cdot \|y\|}, \quad (5.4)$$

kde $c(x, y) := 0$ pokud x nebo y je nula. [14, 22]

5.2.2 Deformační cesta

Uvažujme dvě posloupnosti $X = (x_1, x_2, x_3, \dots, x_N)$ a $Y = (y_1, y_2, y_3, \dots, y_M)$. Pro vytvoření celkového zarovnání prvků posloupností X a Y uvažujme páry indexů, které splňují jisté podmínky. Toto vede na následující definici: An (N, M) - deformační cesta délky $L \in \mathbb{N}$ je posloupnost

$$P = (p_1, \dots, p_L), \quad (5.5)$$

kde $p_\ell = (n_\ell, m_\ell) \in [1 : N] \times [1 : M]$ pro $\ell \in [1 : L]$ splňuje tři následující podmínky:

1. Podmínky ohraničení:

$$p_1 = (1, 1) \text{ a } p_L = (N, M). \quad (5.6)$$

2. Podmínku monotónnosti

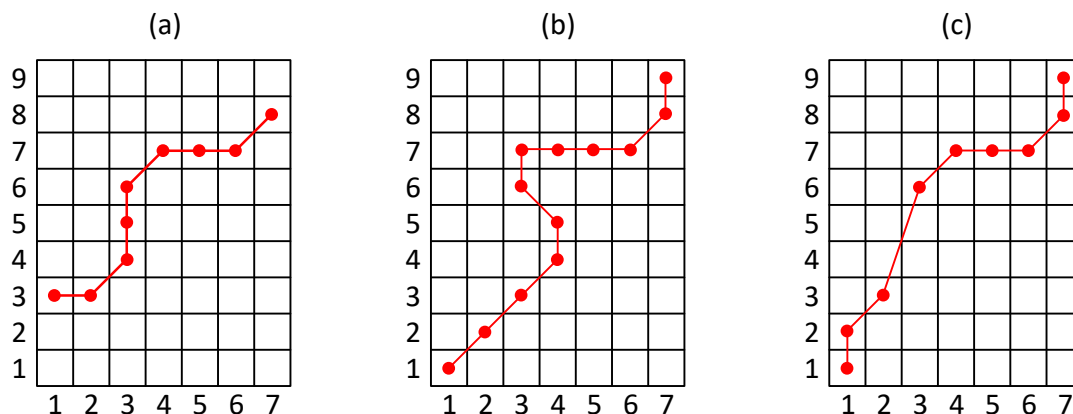
$$n_1 \leq n_2 \leq \dots \leq n_L \text{ a } m_1 \leq m_2 \leq \dots \leq m_L. \quad (5.7)$$

3. Podmínku maximálního kroku

$$p_{\ell+1} - p_\ell \in \{(1, 0), (0, 1), (1, 1)\} \text{ pro } \ell \in [1 : L - 1]. \quad (5.8)$$

An (N, M) - deformační cesta $P = (p_1, \dots, p_L)$ definuje zarovnání dvou sekvencí X a Y přiřazením prvků x_{n_ℓ} posloupnosti X prvkům y_{n_ℓ} posloupnosti Y .

Na obrázku 5.1 můžeme vidět příklady porušení podmínek deformační cesty. V případě a) se jedná o porušení podmínky ohraničení 5.6, b) znázorňuje porušení podmínky monotónnosti 5.7 a c) vyobrazuje porušení podmínky maximálního kroku deformační cesty 5.8. [3, 14, 22]



Obr. 5.1: Zobrazení příkladů porušení podmínek deformační cesty

6 MIR PARAMETRY

V této kapitole budou popsány parametry, které můžeme extrahovat z hudebních nahrávek. Díky těmto parametrům můžeme získat informaci např. o tempu skladby nebo o její harmonii. Tyto parametry mohou být extrahovány různými prostředky. V našem případě budeme využívat prostředí programu *MATLAB*. Konkrétně bude k extrakci využíván volně stažitelný externí toolbox - *MIR Toolbox*. MIR parametry můžeme kategorizovat do několika skupin. Níže budou popsány nejpoužívanější MIR parametry.

6.1 Nízkoúrovňové parametry

Tyto parametry jsou extrahovány přímo z reprezentací hudebního signálu. Všechny níže popsané parametry jsou počítány z datových rámců, jímž odpovídá určitý počet vzorků hudebního signálu. Rozlišujeme nízkoúrovňové parametry oblasti časové a frekvenční.

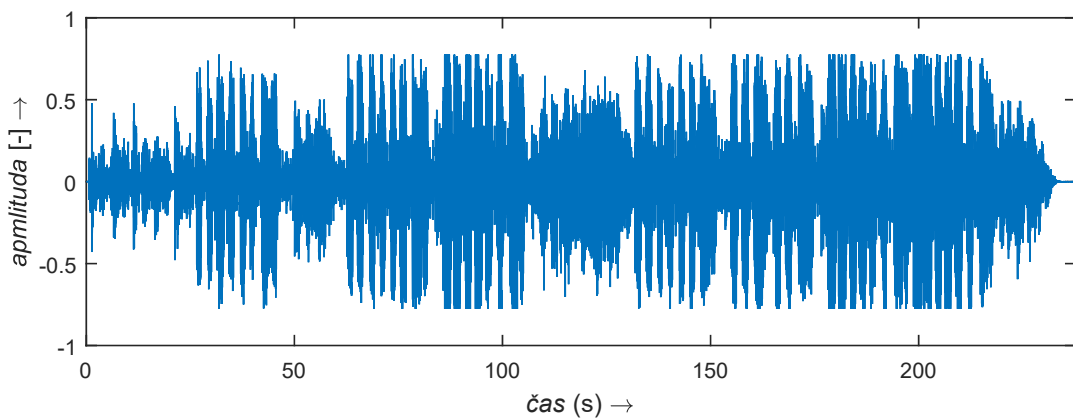
6.1.1 Nízkoúrovňové parametry časové oblasti

Amplitudová obálka

Amplitudová obálka nám udává vizuální informaci o vývoji amplitudy sledovaného signálu v čase t . V podstatě kopíruje špičky vzorků signálu a udává nám tak vnější tvar signálu. Amplitudová obálka je definována vztahem:

$$AE_t = \max_{k=t-K}^{(t+1) \cdot K-1} s(k), \quad (6.1)$$

kde K je velikost rámce vzorků (počet vzorků v každém rámci) a $s(k)$ je amplituda k -tého vzorku. Na obrázku 6.1 můžeme vidět příklad amplitudové obálky.



Obr. 6.1: Amplitudová obálka audio signálu

Efektivní hodnota signálu

Efektivní hodnota signálu (RMS - *Root Mean Square*) jistým způsobem popisuje dynamiku hudebního signálu. Pokud dynamický rozsah hudebního signálu bude malý, bude se její hodnota blížit špičkové hodnotě. Efektivní hodnota je definována vztahem:

$$RMS_t = \sqrt{\frac{1}{K} \cdot \sum_{k=t \cdot K}^{(t+1) \cdot K - 1} s(k)^2}, \quad (6.2)$$

kde K je velikost rámce vzorků (počet vzorků v každém rámci) a $s(k)^2$ je druhá mocnina amplitudy k -tého vzorku. [8]

Počet průchodů nulou

Parametr počet průchodů nulo označován také jako ZRC - *Zero crossing rate* udává kolikrát se změnilo znaménko hodnoty amplitudy. Využívá se při detekci perkusivních zvuků. Je definována vztahem:

$$ZRC_t = \frac{1}{2} \cdot \sum_{k=t \cdot K}^{(t+1) \cdot K - 1} |\text{sgn}(s(k)) - \text{sgn}(s(k+1))|, \quad (6.3)$$

kde K je velikost rámce vzorků (počet vzorků v každém rámci) a $s(k)$ je amplituda k -tého vzorku. [8]

6.1.2 Nízkoúrovňové parametry frekvenční oblasti

Poměr energií pásem

Poměr energií pásem nebo-li BER - *Band Energy Ratio* je používán při klasifikaci hudebních žánrů. Pomocí tohoto parametru lze také rozlišit obsah analyzovaného audio signálu na úrovni mluvené slovo/hudba. Parametr BER udává poměr energií nízkých pásem spektra k energii pásmům vyšších kmitočtů. Tento parametr je definována vztahem:

$$BER_t = \frac{\sum_{n=1}^{F-1} m_t(n)^2}{\sum_{n=F}^N m_t(n)^2}, \quad (6.4)$$

kde F je frekvenční pásmo, které dělí pásmo nízkých kmitočtů od vyšších. Volba F je kritická pro výslednou hodnotu BER, protože ve většině případů energie v nízkých pásmech je větší než v pásmech vyšších. [8]

Šířka pásma

Tento parametr je odvozen z parametru spektrální centroid, který je popsán v kapitole 6.2.4 a udává šířku pásma tedy oblast spektra, ve které se nacházejí nejdůležitější

kmitočty v okolí pásma s nejvyšší energií. Průměrná šířka pásma může být sloužit k popisu barvy zvuku hudební nahrávky. Šířka pásma je definována vztahem:

$$BW_t = \frac{\sum_{n=1}^N |n - SC_t| \cdot m_t(n)}{\sum_{n=1}^N m_t(n)}, \quad (6.5)$$

kde N je číslo nejvyššího kmitočtu a $m_t(n)$ je modul signálu ve frekvenční oblasti za čas t , ve frekvenčním pásmu n . U hudby klasické je šířka pásma menší než u hudebních žánrů typu metal nebo u elektronické hudby. [8]

Spektrální fluktuace

Parametr spektrální fluktuace (SF - *Spectral Flux*) udává změny ve výkonovém spektru mezi po sobě jdoucími rámci. Tento parametr slouží pro detekci počátku not. Je definována vztahem:

$$SF_t = \sum_{n=1}^N (D_t(n) - D_{t-1}(n))^2, \quad (6.6)$$

kde D_t je rámec po rámci normalizovaná frekvenční distribuce za čas t a N je číslo nejvyššího kmitočtu. [8]

6.2 Parametry popisující barvu zvuku

6.2.1 Parametry popisující křivku počátků zvukových událostí

Doba náběhu

Doba náběhu označována také jako *attack time* je parametr, který může být získán např. z amplitudové obálky jednotlivých událostí hudebního signálu. Je definován jako čas za který se změní hodnota energie události z 10% na hodnotu 90% celkové energie jednotlivé události. [10, 16]

Sklon náběhu

Sklon náběhu (*attack slope*) je definován jako průměrný nárůst energie během fáze náběhu. Nejprve jsou spočítány lokální příspěvky energie a po té je spočítán vážený průměr. Váhy jsou vybírány tak, aby zdůraznily hodnotu nárůstů v polovině fáze náběhu. [17]

Skok náběhu

Parametr skok náběhu (*attack leap*) udává rozdíl amplitud maximální hodnoty ve fázi náběhu a lokálního minima před touto špičkou.

Sklon poklesu

Výpočet sklonu poklesu (*decay slope*) je založen na následujícím modelu začínajícím v bodě největší energie obálky. Pomocí tohoto parametru lze rozlišit zvuky z hlediska jejich délky, tedy lze rozlišit zvuky perkusivní od zvuků soustavných. Sklon poklesu je definována vztahem:

$$S(t) = A \cdot \exp(-\alpha(t - t_{max})) \cdot t_{nmax}, \quad (6.7)$$

kde veličina α je odhadnuta lineární regresí logaritmu obálky energie. [29, 17]

Doba trvání

Doba trvání (*duration time*) udává dobu za kterou obálka energie přesáhne jistou prahovou úroveň, obvykle 40%. Tímto parametrem lze určit např. dobu trvání noty v sekundách nebo rozlišit zvuky perkusivní od zvuků soustavných. [10]

6.2.2 Parametry popisující rozložení energie ve spektru

Pokles spektrální energie

Pokles spektrální energie (*spectral rolloff*) udává hodnotu frekvence, pod kterou se na nachází určité množství energie. Obvykle je toto množství energie vyjádřeno hodnotou 85%. Tento parametr je počítán ze spektra analyzovaného signálu a slouží k odhadnutí množství vysokých frekvencí v signálu. [10]

Bělost spektra

Tento parametr slouží k určení množství energie (vyjádřeno v procentech), která se nachází nad jistou frekvencí. Běžně se volí kmitočty o hodnotě 1 kHz, 1.5 kHz nebo 3 kHz. Bělost spektra bývá v anglicky psané literatuře označována jako *spectral brightness*. [10]

6.2.3 Drsnost spektra

Parametr drsnosti spektra (*spectral roughness*) byl navrhnut v roce 1965 pro odhad disonance (nelibovosti) zvuku. Ta vzniká pokud jsou dvě sousedící harmonické složky příliš blízko sebe. Tedy čím více harmonických složek s dostatečně velkou amplitudou sousedí blízko sebe, tím větší je drsnost (disonance) zvuku. V praxi je tento parametr počítán ve třech krocích. Nejprve je spočítáno spektrum analyzovaného signálu a z toho jsou vypočítány lokální maxima spektra. Až v posledním kroku je počítána drsnost spektra jako poměr amplitud všech možných sousedících maxim.

6.2.4 Parametry spektrálního rozložení

Tyto parametry popisují prakticky tvar spektra a jeho rozložení pomocí statistických momentů.

Spektrální centroid

Spektrální centroid (SC - *Spectral Centroid*) udává frekvenční pásmo, ve kterém je koncentrováno nejvíce energie analyzovaného signálu. Pomocí tohoto parametru je měřena tzv. „jasnost“ barvy zvuku. Parametr SC je velice citlivý na filtraci signálu filtrem, který upravuje velikost vyšších kmitočtů např. filtru typu dolní propust, protože dává větší váhu kmitočtům ve vyšších pásmech. Obecně tento problém také nastane při operaci pod vzorkování signálu, protože tato operace zapříčiňuje ořez vyšších kmitočtů signálu. U hudby klasické jsou hodnoty parametru SC mnohem nižší a proměnlivost menší, než u metalu nebo elektronické hudby. Parametr SC je definována vztahem:

$$SC_t = \frac{\sum_{n=1}^N m_t(n) \cdot n}{\sum_{n=1}^N m_t(n)}, \quad (6.8)$$

kde N je číslo nejvyššího kmitočtu a $m_t(n)$ je modul signálu ve frekvenční oblasti za čas t , ve frekvenčním pásmu n . [8]

Spektrální entropie

Pojem entropie jako takový, představuje (zjednodušeně řečeno) míru neurčitosti. Normální (Gaussovo) rozložení má nejvyšší míru entropie.

Spektrální plochost

Parametr plochost spektra (*spectral flatness*) popisuje tvar spektra. Je vypočítán jako poměr geometrického průměru spektra a aritmetického průměru. Hodnota plochosti spektra bílého šumu se blíží jedné ze zdola, protože bílý šum má rovnoměrné rozložení výkonové spektrální hustoty. Naopak čím více bude je spektrum signálu členité tím více se hodnota plochosti spektra se bude blížit nule ze shora. Plochost spektra je definována vztahem:

$$F_t = \frac{\sqrt[N]{\prod_{n=1}^N m_t(n)}}{\frac{\sum_{n=1}^N m_t(n)}{N}}, \quad (6.9)$$

kde N je číslo nejvyššího kmitočtu a $m_t(n)$ je modul signálu ve frekvenční oblasti za čas t , ve frekvenčním pásmu n . [8]

Spektrální špičatost

Spektrální špičatost (*spectral kurtosis*) je parametr, který udává míru „špičatosti“ spektra v okolí jeho střední hodnoty. Normální rozdělení má hodnotu koeficientu špičatosti rovno nule. Kladná hodnota koeficientu špičatosti značí, že většina spektrálních složek leží blízko střední hodnoty a tudíž je spektrum špičatější. Naopak pokud je hodnota záporná, spektrum je plošší a rozložení spektrálních složek je rovnoměrnější. [10]

Spektrální šikmost

Tento parametr udává míru asymetrie spektra okolo jeho střední hodnoty. Pokud je hodnota koeficientu šikmosti rovna nule, jedná se o symetrické rozložení spektrálních složek kolem střední hodnoty spektra. Jestliže je hodnota větší jak nula je více spektrálních složek zastoupeno ve vysokých kmitočtech a naopak, pokud je hodnota koeficientu šikmosti spektra menší jak nula, je více spektrálních složek obsaženo v kmitočtech nízkých. Původní název toho parametru je *spectral skewness*. [10]

6.2.5 Melovské keprální koeficienty

Melovské keprální koeficienty (*MFCC*) jsou nejvíce používané v systémech automatického rozpoznávání řeči, ale v poslední době našly uplatnění také v MIR. Jejich výpočet se skládá z pěti kroků:

1. Analyzovaný signál je rozdělen do rámců dlouhých 20–40 ms a ořezání okrajů okénkem, nejčastěji Hammingovým
2. Fourierova transformace signálu
3. Rozdělení spektra bankou 40 trojúhelníkových filtrů a následný přepočít na MEL-frekvenci
4. Výpočet absolutní hodnoty, logaritmování
5. Výpočet diskrétní kosinové transformace

MEL-frekvence je škála frekvence, která nejlépe popisuje reálné lidské vnímání výšky tónu. Tato škála je cca do $f = 1$ kHz lineární a nad tuto frekvenci je logaritmická. Přepočít je definován vztahem:

$$Mel(f) = 1125 \cdot \log \left(1 + \frac{f}{700} \right). \quad (6.10)$$

Melovské keprální koeficienty jsou v MIR používány pro klasifikaci hudebních žánrů nebo při rozpoznávání hudebních emocí. [29]

6.3 Parametry popisující dynamiku hudební nahrávky

6.3.1 Efektivní hodnota signálu

Tento parametr byl popsán v kapitole 6.1.1.

6.3.2 Segmentace signálu

Segmentace je provedena na pozicích dlouhého ticha. Signál je rozdělen na rámce. Pro každý rámec je vypočítána efektivní hodnota signálu. Segmenty jsou poté vybrány z časových pozic, ve kterých efektivní hodnota je větší než definovaná prahová úroveň, do úrovně kde efektivní hodnota klesne pod prahovou úroveň. [10]

6.3.3 Parametr nízké energie

Křivka nízké energie (*low energy*) může být použita k vyhodnocení časového rozložení energie. Lze pomocí ní určit zda energie zůstává konstantní v celém pásmu nebo zda jsou některé úseky více kontrastní než jiné. Jeden ze způsobů výpočtu hodnoty nízké energie je výpočet míry nízké energie, tedy počet rámců, které mají energii menší než je průměrná energie. [10]

6.4 Parametry popisující tempo hudební nahrávky

6.4.1 Detekce počátku zvukové události

Metoda detekce počátku zvukové události je založena na principu trazientních signálů, které jsou typické rychlou změnou kmitočtu a nacházejí se ve fázi náběhu tónu. Tato metoda je prováděna obecně v těchto krocích:

1. Signál je rozdělen na rámce malé velikosti.
2. Je provedeno předzpracování dat (volitelné).
3. Je proveden výpočet detekce počátku zvukové události pomocí vhodné funkce.
4. Výsledek výpočtu z předcházejícího kroku je normalizován.
5. Normalizovaný výsledek je prahován.
6. Jsou lokalizovány počátky zvukových událostí.

Pomocí této metody jsem tedy schopni určit polohu tranzientních signálů. Díky tomu jsem schopni odhadnout tempo analyzované hudební nahrávky. [29]

6.4.2 Hustota zvukových událostí

Jedná se o parametr, který udává průměrnou frekvenci počtu zvukových událostí za jednu sekundu. Tento parametr bývá často označován pojmem *tempo density*. [10]

6.4.3 Fluktuace tempa

Tento parametr je počítán ze spektrogramu signálu, který byl rozdělen na segmenty o délce okna 23 ms. Poté je frekvenční osa přepočítána na melovskou či barkovou škálu. Následně jsou zahrnuty do výpočtů maskovací efekty lidského ucha. Posledním krokem je výpočet FFT každého segmentu. Výstupem je matice rytmické periodicity.

6.4.4 Tempo

Tempo je definováno jako počet úderů za časový úsek. Jak bylo zmíněno v kapitole 4.3 jeho hodnota je uvedena v jednotkách BPM. Výpočet tohoto parametru je prováděn pomocí detekcí periodicit z křivky počátků zvukových událostí. [10]

6.4.5 Metrický centroid

Tímto parametrem může být posouzena metrická aktivita na základě výpočtu centroidu vybrané metrické úrovně. Výsledná křivka metrického centroidu vypovídá o vývoji metrické aktivity vyjádřené v jednotkách BPM. Díky tomu může být výsledek přímo porovnán s hodnotami tempa v BPM. Velmi rychlé metrické úrovně odpovídají rychlým rytmickým hodnotám a naopak nízké pulzace odpovídají notám celým. [10]

6.4.6 Metrická síla

Metrická síla naznačuje vývoj tempa v čase. Vysoké hodnoty obvykle znázorňují více elementární rytmické dělení skladby a naopak nízké hodnoty ukazují složitější rytmické dělení. [10]

7 STATISTICKÉ PARAMETRY

7.1 Střední hodnota

Střední hodnota (anglicky *mean*) je nejznámější míra polohy ve statistice a značí se $EX, E(X)$. Je to statický parametr rozdělení náhodné veličiny, který je definován jako vážený průměr daného rozdělení [6]. Pro náhodnou veličinu X s diskrétním rozdělením je definována:

$$E(X) = \sum_I s_i p_i, \quad (7.1)$$

kde $P[X = s_i] = p_i$ pro $i \in I$.

7.2 Směrodatná odchylka

Jedná se o kvadratický průměr odchylek hodnot znaku od jejich aritmetického průměru. Je-li její hodnota malá, jsou si prvky souboru většinou podobné a naopak. V matematice je označována řeckým písmenem σ , ale v prostředí *MATLAB* je označena zkratkou *std* [6]. Její hodnota lze vypočítat ze střední hodnoty $E(X)$:

$$\sigma = \sqrt{E((X - E(X))^2)}. \quad (7.2)$$

7.3 Rozptyl

Rozptyl je střední hodnota kvadrátů odchylek od střední hodnoty $E(X)$, vyjadřuje tedy variabilitu rozdělení souboru náhodných hodnot kolem její střední hodnoty [6]. Bývá označován σ^2 nebo *var*. Je definován vztahem:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - E(X))^2. \quad (7.3)$$

7.4 Kvantily

Kvantily jsou hodnoty, které dělí soubor seřazených hodnot na několik zhruba stejně velkých částí. Kvantil označujeme písmenem Q s dolním indexem, označující hodnotu rozdělení souboru hodnot. [6]

7.4.1 Medián

Medián je kvantil rozdělující statistický soubor na dvě stejně početné množiny. Bývá označován $Q_{0.5}$.

7.4.2 Kvartil

Tři kvartily rozdělují statistický soubor na čtvrtiny. Jako dolní kvartil - $Q_{0.25}$ je označováno 25% prvků, jako horní kvartil - $Q_{0.75}$ je označováno 75% prvků. Dolní/horní kvartil bývá také označován jako Q_1/Q_3 .

7.4.3 Mezikvartilové rozpětí

IQR (*Interquartile range* - mezikvartilové rozpětí) představuje rozdíl mezi třetím a prvním kvantilem. Reprezentuje tedy oblast hodnot, které má střeňích 50% hodnot proměnné. Je definováno jako:

$$IQR = Q_3 - Q_1. \quad (7.4)$$

7.4.4 Percentil

Percentil rozděljuje statistický soubor na setiny. K-tý percentil označujeme $Q_{\frac{k}{100}}$.

7.4.5 Mezipercentilové rozpětí

IR (*Interpercentile range* - mezipercentilové rozpětí) udává rozdíl mezi devadesátým devátým percentilem a jedním percentilem. Je definován jako:

$$IR = Q_{0.99} - Q_{0.01}. \quad (7.5)$$

8 DATABÁZE INTERPRETACÍ

V této kapitole bude popsáno jakým způsobem byla vytvořena databáze interpretací, která bude sloužit jako vstupní data pro vyhodnocovací systém. Z důvodů výpočetní náročnosti nebudou testovány skladby celé, pouze jejich dílčí úseky. Tyto úseky byly vybrány na základě instrumentační analýzy skladby, která bude popsána v podkapitole 8.1. Pro analýzu byla vybrána skladba od skladatele Ludwiga von Beethovena, konkrétně jeho symfonie číslo 5, známá také jako „osudová“. Notový zápis celé této první věty bude obsažen na přiloženém CD. Důvodem výběru této skladby byl fakt, že se dle mého názoru jedná o nejznámější symfonii vůbec a tím pádem nebyla nouze nalézt nespočet různých interpretací.

8.1 Rozbor skladby

Tato kapitola je věnována rozboru skladby od skladatele Ludwiga von Beethovena, konkrétně první věty jeho páté symfonie, která byla vybrána pro analýzu. Tato symfonie poprvé zazněla 22. prosince 1808 na Beethovenově akademii ve vídeňském *Theater an der Wien*. Její celý název zní: *Symfonie č. 5 C moll op. 67*. Celková délka této symfonie je přibližně 32 minut a skládá se ze čtyř částí:

1. *Allegro con brio (C moll)*.
2. *Andante con moto (As moll)*.
3. *Scherzo: Allegro (C moll)*.
4. *Allegro (C dur)*.

I když se v průběhu času vyskytly instrumentační úpravy této symfonie pro různé nástroje např. pro sólový klavír, původně je instrumentační obsazení této symfonie je následující:

- Smyčcové nástroje: 1. a 2. housle, violy, violoncella, kontrabasy.
- Dřevěné dechové nástroje: 2 flétny, pikola, 2 hoboje, 2 klarinety, 2 fagoty a kontrafagot.
- Žestové nástroje: 2 trubky, 2 lesní rohy a 3 trombóny.
- Bicí nástroje: tympány.

8.1.1 První věta (*Allegro con brio*)

První věta páté symfonie je napsána formou sonáty¹. Tuto formu si Beethoven osvojil od svých klasicistních učitelů Haydna a Mozarta. První věta začíná dvěma frázemi

¹V této formě se obecně na prvních stránkách objevují hlavní myšlenky (motivy) skladby, později procházejí vývojem (modulací do různých tónin) a zhruba ve třech čtvrtinách skladby je dramatický návrat k hlavním motivům skladby, které se objevily na začátku.

hrané v dynamice *fortissimo*, v níž zazní slavný motiv „ta-ta-ta- taaa“, který doslova ovládne posluchačovu pozornost. V následujících čtyřech taktech Beethoven používá imitace a sekvence k rozšíření hlavního motivu. Tyto imitace se překrývají navzájem s takovou rytmickou pravidelností, že se zdá, že tvoří jedinou celistvou melodii. Krátce poté zazní před druhým tématem velmi krátký *fortissimo* přechod, který hrají lesní rohy. Druhé téma moduluje do tóniny *Es dur* a působí tak více lyricky. Následuje *coda*, které je opět založeno na hlavním motivu. V průběhu repetice zazní krátké hobojové sólo a věta končí epickým vyvrcholením.

8.2 Databáze

Jako zdroj dat byly použity zvukové nahrávky první věty (*Allegro con brio*) Beethovenovy páté symfonie v podání symfonického orchestru. Tyto nahrávky byly získány v bezztrátových formátech (.m4a, .flac apod.), které byly sjednoceny převodem do formátu .wav s parametry CD standardu.

Databáze je tvořena 56 verzemi této skladby v podání symfonických orchestrů z celého světa, se širokou škálou data nahrání. Seznam nahrávek je uveden v příloze B. Tyto hudební nahrávky budou kategorizovány do tří tříd:

- interpretace nahrány v období 1950–1969 (celkem 18 prvků),
- interpretace nahrány v období 1970–1989 (celkem 18 prvků),
- interpretace nahrány v období 1990–dnes (celkem 20 prvků).

Tato databáze bude sloužit jako vstupní data pro vyhodnocovací systém, který je popsán v kapitole 9. Pro dosažení optimálních výsledků vyhodnocovacího systému by tato databáze měla mít konzistentní počet prvků v každé kategorii a zároveň by počet prvků v kategoriích měl být co nejvyšší.

Při vytváření této databáze byl opravdu velký problém získat relevantní informace o tom, kdy byla daná nahrávka natočena. Tato informace je pro vyhodnocovací systém klíčová. Slouží totiž jako referenční parametr. Špatné určení data nahrání a tedy i následné mylné zařazení dané nahrávky do nesprávné kategorie, může způsobit velké zkreslení konečných výsledků a dojde ke „zmatení“ celého vyhodnocovacího systému. Data nahrání jednotlivých nahrávek byly určeny pomocí internetové hudební encyklopedie – *Music Brainz* [11].

8.3 Výběr úseků k analýze

Díky výpočetní náročnosti výpočtů nebudou analyzovány zvukové nahrávky celé, nýbrž jejich dílčí úseky. Tím bude zároveň možné si vytvořit představu o tom, které části skladeb jsou vhodnější pro takovouto analýzu. Délka vybraných úseků by měla

být kompromisem mezi přílišnou délkou úseku (kvůli výpočetní náročnosti) a nedostačující délkou (příliš málo obsažených informací).

V této práci budou analýze podrobeny dva úseky, které byly vybrány na základě instrumentálního rozboru skladby. Bylo také cílem vybrat co možná nejkontrastnější úseky. Proto byl při výběru brán také zřetel na dynamiku hry, ale i na barvu zvuku (spektrum) úseku. Nutno podotknout, že byla snaha také splnit kritérium délky úseku. To bylo stanoveno, dle návrhu vedoucího práce, na minimální délku 30 sekund a maximální délku úseku 60 sekund.

První úsek

Tento úsek je dlouhý 25 taktů. Jedná se o část skladby, která je v dynamice *piano* s gradací do *forte fortissima*. Výrazným prvkem toho úseku je krátké hobojové sólo. Tento úsek se nachází v poslední části skladby mezi repeticí hlavního motivu, před epickým vyvrcholením.

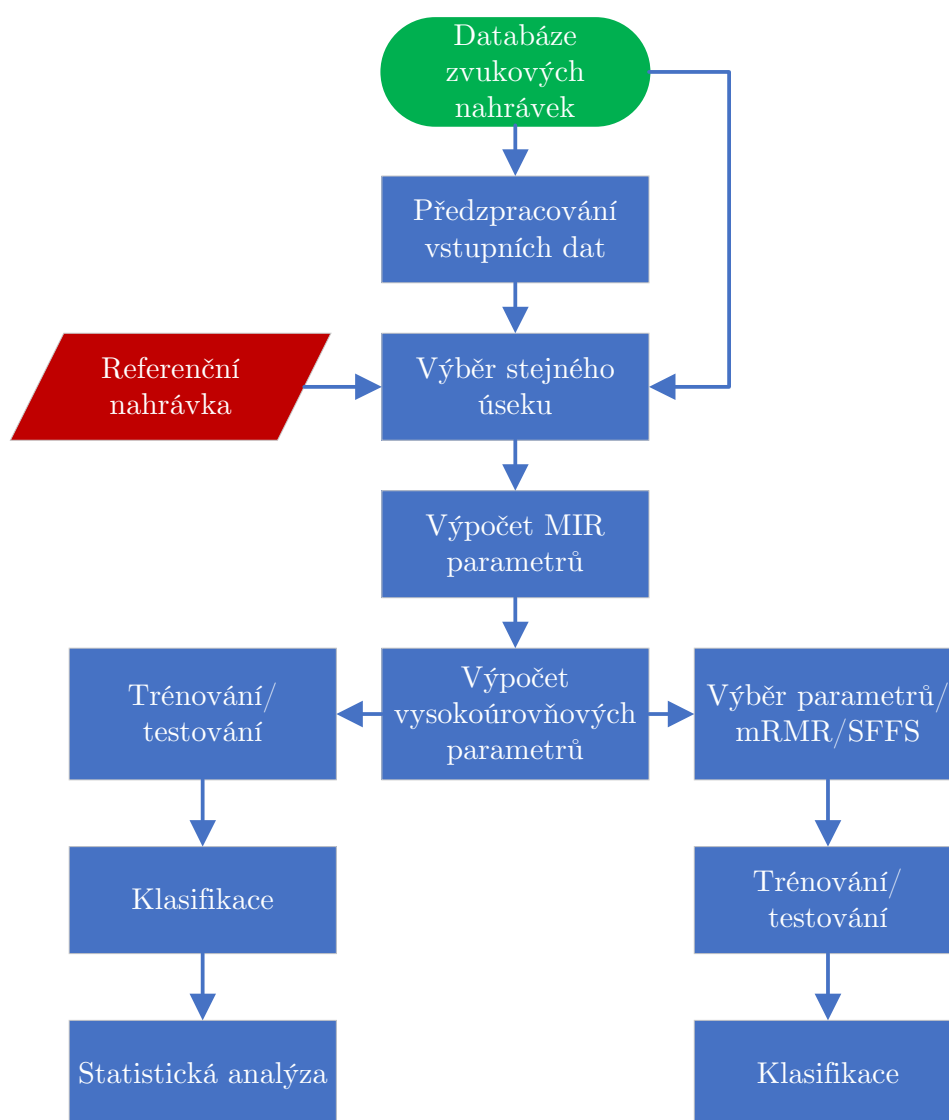
Druhý úsek

Druhý vybraný úsek je kontrastním prvkem úseku prvního. Celý tento úsek je zapsán v dynamice *forte fortissimo*. Nachází se prakticky na konci skladby a tvoří epické vyvrcholení. Tento úsek je dlouhý 70 taktů.

Notový zápis úseku prvního je obsažen v příloze C a úseku druhého v příloze D. Začátek vybraného úseku je v notovém zápise označen písmenem A a konec písmenem B nad konkrétním taktem.

9 VYHODNOCOVACÍ SYSTÉM

Tato kapitola se zabývá popisem navrženého vyhodnocovacího systému a jeho dílčími bloky. Jako vstupní data slouží vytvořená databáze nahrávek, která byla blíže popsána v kapitole 8. Úkolem toho vyhodnocovacího systému je vyhodnocovat období vzniku zvukové nahrávky. Bylo tedy nutné vypočítat co nejvíce parametrů z vybraných úseků. Dále tyto parametry podrobit individuální analýze a také najít kombinace parametrů, dle kterých lze co nejlépe určit období vzniku zvukové nahrávky. Tento vyhodnocovací systém byl navržen v prostředí programu MATLAB a jeho blokové schéma můžeme vidět na obrázku 9.1.



Obr. 9.1: Blokové schéma vyhodnocovacího systému

9.1 Předzpracování dat

V prvním bloku vyhodnocovacího systému dojde k několika operacím s načteným zvukovým souborem z databáze. Nejprve je provedena normalizace, poté je zvukový soubor převeden ze stereo formátu do mono formátu. Dále je signál z důvodů snížení výpočetní náročnosti DTW (bude provedeno v dalším bloku) pod vzorkován na poloviční vzorkovací kmitočet, tedy 22050 Hz. Tyto operace byly provedeny v Matlabu v jediném kroku pomocí funkce, která je obsažena v MIR Toolboxu `miraudio(X, 'Sampling', 22050, 'Mono', 'Normal')`. Kde `X` je vstupní posloupnost, v tomto případě zvuková nahrávka.

9.2 Výběr stejných úseků

Jak již bylo zmíněno v kapitole 5 délky nahrávek, z důvodů odlišných temp, nejsou stejné. Pro úspěšnou statistickou analýzu je potřeba zajistit synchronizaci všech nahrávek obsažených v databázi a poté zajistit výběr stejného notového úseku. Tento úkon je proveden právě v tomto bloku.

Prvním krokem je převedení předzpracovaných zvukových souborů na posloupnosti, které popisují tónový obsah zvukové nahrávky - chroma posloupnosti. Nejprve je vypočítán parametr s názvem *Pitch*. Ten reprezentuje rozložení jednotlivých tónů v čase. Tento krok je proveden pomocí funkce `audio_to_pitch_via_FB(X)`. Výstupem této funkce je $120 \times N$ matice, kde N je počet segmentů a první dimenze odpovídá výškám tónů dle MIDI.

Dále je parametr *Pitch* přepočítán na parametr *CENS*, čímž je získán vektor o 12 dimenzích. Každá dimenze představuje jednu chroma skupinu. Tento přepočet je zajištěn pomocí funkce `pitch_to_CENS(X)`. Obě funkce nejsou v klasickém Matlabu běžně implementovány a byly získány z volně stažitelného Chroma Toolboxu.

Vypočítaný parametr *CENS* slouží jako vstupní data pro algoritmus DTW, který synchronizuje danou posloupnost s posloupností referenční. Jako referenční byla zvolena nahrávka Berlínského symfonického orchestru z roku 1982 pod vedením světoznámého dirigenta Herberta von Karajana. Důvodem výběru této skladby za referenční je fakt, že její délka prakticky odpovídá mediánu délkám všech ostatních zvukových nahrávek databáze. Výstupem algoritmu DTW jsou dvě proměnné. První se nazývá *cost* neboli cena a je to skalární proměnná. Tato proměnná nabývá hodnot od 0 do 1 a vyjadřuje míru podobnosti synchronizovaných posloupností. Čím více se hodnota proměnné *cost* blíží k nule tím více jsou si posloupnosti podobné. Druhou výstupní proměnnou je tzv. *wrapping path*. Tato proměnná je vyjádřena formou dvousloupcového vektoru a obsahuje informaci o optimální cestě.

Posledním krokem je samotný stříh vybraného úseku. Ten byl proveden funkcí z MIR Toolboxu `miraudio(X, 'Extract', t1, t2)`, kde X je vstupní nahrávka, $t1$ je čas začátku stříhu a $t2$ je čas konce stříhu. Časy stříhů pro každou nahrávku z databáze byly odečteny z vypočítaných proměnných *wrapping path*.

Celkem byly vybrány dva kontrastní úseky (viz. kapitola 8.3). Jelikož databáze čítá 56 prvků, počet souborů pro další zpracování je roven dvojnásobku počtu prvků databáze, tedy 112. Nutno podotknout, že vybrané úseky byly stříhány z původních zvukových nahrávek, tedy nedošlo k žádné ztrátě informace.

9.3 Výpočet parametrů

V tomto bloku dochází k výpočtu parametrů z vybraných úseků. Ty byly vypočítány pomocí funkcí z již zmiňovaného MIR Toolboxu. Cílem bylo vypočítat co nejvíce parametrů, které pomohou správně klasifikovat dané nahrávky do správných kategorií. Tyto parametry byly rozděleny do 5 následujících kategorií a jejich popis je možné dohledat v kapitole 6:

1. parametry popisující rozložení spektrální energie
 - bělost spektra nad $f = 1$ kHz (*bright 1k*)
 - bělost spektra nad $f = 1,5$ kHz (*bright 1,5k*)
 - bělost spektra nad $f = 3$ kHz (*bright 3k*)
 - pokles spektrální energie 85% (*roll of 85*)
 - pokles spektrální energie 95% (*roll of 95*)
2. parametry popisující spektrální rozložení
 - spektrální centroid (*spectral centroid*)
 - spektrální entropie (*spectral entropy*)
 - spektrální plochost (*spectral flatness*)
 - spektrální špičatost (*spectral kurtosis*)
 - spektrální šikmost (*spectral skewness*)
 - spektrální rozptyl (*spectral spread*)
3. parametry popisující křivku zvukových událostí
 - doba náběhu (*attack time*)
 - strmost náběhu (*attack slope*)
 - skok náběhu (*attack leap*)
 - strmost odeznění (*decay slope*)
4. Melovské keprstrální koeficienty 2-13
5. Ostatní
 - efektivní hodnota signálu (*RMS*)
 - parametr nízké energie (*low energy*)
 - počet průchodů nulovou úrovní (*ZRC*)

- drsnost spektra (*spectral roughness*)
- nepravidelnost spektra (*spectral irregularity*)
- tempo
- fluktuace tempa (*fluctuation*)

Každý takto vypočítaný parametry je reprezentován formou vektoru. Proto z každého toto parametru byly vypočítány vysokoúrovňové statické parametry (viz. tabulka 9.1) a ty slouží jako vstupní data pro klasifikaci samotnou.

Tab. 9.1: Přehled statistických parametrů

Zkratka parametru	Popis
mean	střední hodnota
var	rozptyl
std	směrodatná odchylka
median	medián
1p	1. percentil
99p	99. percentil
ir	mezipercentilové rozpětí
q1	1. kvartil
q3	3. kvartil
iqr	mezikvartilové rozpětí

Celkem bylo tedy vypočítáno přesně 340 parametrů. Jejich schopnost diferencovat jednotlivá časová období bude více popsána v kapitole 10.

9.4 Výběr parametrů

I když bylo cílem vypočítat co nejvíce parametrů, nemusí tato skutečnost přispět k nejlepším výsledkům samoučících se algoritmů. Mohou nastat komplikace ve formě velké výpočetní náročnosti, vedoucí k dlouhé době výpočtů, malé efektivitě a také může dojít k přetrénování klasifikátoru. Z tohoto důvodu je vhodné odstranit nepotřebné parametry a také parametry, které jsou vzájemně korelovatelné. Pro tento proces selekce parametrů byly vytvořeny různé metody. Obecně je označujeme jako metody výběru optimálních parametrů a dělíme je do dvou skupin:

- metody optimalizace bez použití klasifikátoru. Ty označujeme jako metody typu „filtering“. Nejpoužívanější metodou této skupiny je metoda minimální redundance a maximální relevance neboli mRMR (*minimum Redundancy Maximum Relevance* - metoda minimální redundance, maximální relevance) - více v kapitole 9.4.1.

- metody využívající klasifikátor. Tyto metody označujeme jako metody typu „wrapping“. Typickým představitelem těchto metod je metoda plovoucí dopředného výběru SFFS (*Sequential Forward Floating Search*). Více informací o této metodě je uvedeno v kapitole 9.4.2.

Pro zlepšení výsledků klasifikace je v této práci použita metoda selekce parametrů SFFS. Ta hledá kombinace parametrů pro zlepšení přesnosti klasifikace. Výsledná úspěšnost kombinací parametrů je uvedena v kapitole 10.2. [9, 26]

9.4.1 Metoda mRMR

Jak již bylo zmíněno výše, metodu mRMR řadíme mezi metody typu „filtering“. Ty vybírají podmnožinu parametrů nezávisle na použitém klasifikátoru. Její princip by se dal popsat těmito třemi kroky:

1. výběr parametru, který nejvíce souvisí s danou třídou
2. výběr dalšího parametru, jehož rozdíl souvislosti s danou třídou a průměrné hodnoty souvislosti s již vybranými parametry je co nejvyšší
3. opakování, dokud není vybrán požadovaný počet optimálních parametrů.

Tato metoda je význačná její rychlostí a univerzálností, ale nutno podotknout, že poskytuje horší výsledky. Výstupem této metody je seznam parametrů seřazený dle jejich významu. Následně je vybráno pouze n prvních parametrů. [18, 26]

9.4.2 Metoda SFFS

Dále je výběr parametrů proveden metodou sekvenční dopředné selekce (SFFS). Jedná se o metodu výběru optimálního počtu parametrů typu „wrapping“, to znamená, že do výpočtů je zahrnut i klasifikátor. Využívá plovoucího výběru, což znamená, že umožňuje kromě přidávání parametrů i jejich odebírání z již zvolené množiny vybraných parametrů. Tato metoda hledá nejen samostatné parametry, které vykazují vysokou míru úspěšnosti klasifikace, ale také hledá jejich kombinace pro dosažení ještě lepších výsledků klasifikace [19]. Díky tomu je tato metoda velmi přesná, ale na druhou stranu je výpočetně náročná. Její princip lze **zjednodušeně** popsat těmito kroky:

1. výběr parametru, u kterého klasifikátor dosahuje nejvyšší úspěšnosti klasifikace
2. k tomuto parametru je přidán parametr další. Je pozorována úspěšnost kombinace, která poskytuje nejvyšší úspěšnost klasifikace.
3. jsou přidány postupně další parametry, dokud roste úspěšnost klasifikace [26].

Tato metoda je v systému realizována funkcí `SFFS_cls(X,Y,sett)`. Jako předloha sloužila stejnojmenná funkce z volně dostupného softwaru pro sekvenční selekci příznaků [12].

9.5 Klasifikace

Tato kapitola se zabývá popisem klasifikačních metod použitých ve vyhodnocovacím systému. Pomocí těchto metod je možné provést individuální klasifikaci jednotlivých parametrů.

9.5.1 Křížová validace

Jedná se o statistickou metodu, pomocí které jsme schopni určit jako moc bude model statistické analýzy ovlivňovat nezávislé vzory dat. Principem je rozdělení vstupní množiny dat na podmnožiny. Větší část rozdělených dat je použita pro trénování a zbylá menší část vstupních dat pro testování. Klasifikátor poté natrénuje model na testovací podmnožině dat a následně jeho přesnost ověří na podmnožině testovacích dat. Proces je několikrát opakován pokaždé s odlišnou podmnožinou testovacích a trénovacích dat.

Existuje několik metod křížové validace např.: *K-fold cross validation*, *Repeated random subsampling* a *Leave-one-out cross validation*.

V této práci bude použita metoda *Leave-one-out cross validation*. Ta je založena na principu použití jednoho vzorku z celé množiny vstupních dat na testování a zbytek je použit na trénování modelu.

9.5.2 Klasifikační stromy

Klasifikační stromy řadíme mezi neparametrické metody strojového učení. Každý klasifikační strom lze popsat stromovým grafem, který je tvořen uzly a orientovanými hranami. V každém neterminálním uzlu se strom větví. Z uzlu vedou hrany do dvou či více dceřinných uzlů. Větvení stromu je založeno na hodnotě jediného prediktoru. Nejběžnější je tzv. binární větvení dle odpovědi na otázku ve tvaru „ $x_i < c$ “. Poté je kladné odpovědi přiřazena jedna hrana a záporné odpovědi hrana druhá. Testování prediktorů začíná v kořenovém uzlu přes větvení neterminálních uzlech k některému terminálnímu uzlu - listu.

Kombinací více klasifikačních stromů vznikne struktura, kterou označujeme jako klasifikační les. Výsledná klasifikační funkce je vytvořena vhodnou kombinací klasifikačních funkcí jednotlivých stromů. [7]

9.6 Statistická analýza

V této kapitole budou popsány jednotlivé statistické analýzy, které byly ve vyhodnocovacím systému použity. Jejich cílem je analyzovat jednotlivé vypočítané parametry

a určit, pomocí kterých parametrů lze nejlépe určit období nahrání dané interpretace. Níže popsaná metoda - Spearmanův test pořadové korelace patří mezi tak zvané neparametrické metody. Ty se používají tehdy, pokud není možné předpokládat, že náhodné veličiny X, Y mají normální rozdělení.

9.6.1 Přesnost

Jedná se o klasifikátor, který bývá označován také jako *accuracy*. Je to poměr mezi počtem správně klasifikovaných prvků a celkovým počtem prvků. Lze vypočítat z takzvané matice záměn (*confusion matrix*). V případě, že jsou známy pouze dvě klasifikační třídy, operuje se čtyřmi základními možnými výsledky:

- *True positive (TP)*: počet správných pozitivních nálezů
- *False positive (FP)*: počet falešných pozitivních nálezů
- *True negative (TN)*: počet správných negativních nálezů
- *False negative (FN)*: počet falešných negativních nálezů

Celková přesnost je poté definována vztahem:

$$ACC = \frac{TP + TN}{TP + TN + FN + FP} \cdot 100 \quad [\%] \quad (9.1)$$

9.6.2 Spearmanův test pořadové korelace

Opět se jedná o neparametrickou metodu. Její princip je založen na výpočtu pořadí hodnot sledovaných veličin. Spearmanův koeficient pořadové korelace je definován vztahem:

$$\rho = 1 - \frac{6 \sum_{i=1}^n (R_i - Q_i)^2}{n(n^2 - 1)}, \quad (9.2)$$

kde R_i označuje pořadí náhodné veličiny X_i a Q_i udává pořadí náhodné veličiny Y_i . Parametry n určuje počet korelačních dvojic. Koeficient může nabývat hodnot $-1 \leq \rho \leq 1$. Čím blíže je hodnota koeficientu jedné, tím větší je přímá pořadová hodnota mezi náhodnými veličinami X a Y . Pokud naopak je hodnota koeficientu blíže mínus jedné tím silnější je nepřímá závislost. [1]

V Matlabu se výpočet Spearmanova koeficientu pořadové korelace realizuje pomocí funkce: `rho = corr(X,Y,'type','Spearman');`

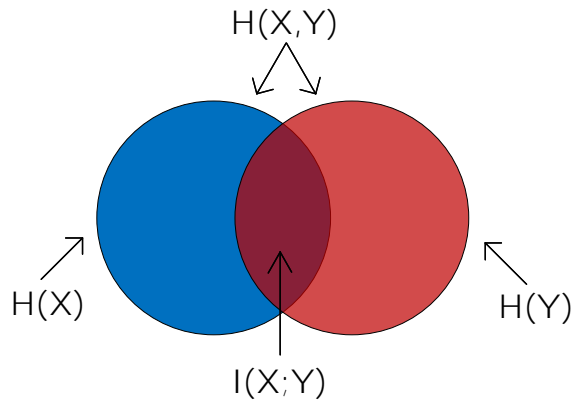
9.6.3 Vzájemná informace

jedná se o veličinu, díky které můžeme porovnat vztah mezi dvěma a více náhodnými veličinami. V případě, že je vzájemná informace nulová, jsou veličiny statisticky

nezávislé. Vzájemnou informaci mezi dvěma náhodnými veličinami můžeme vyjádřit jako:

$$I(X;Y) = H(X) + H(Y) - H(X,Y), \quad (9.3)$$

kde $H(X)$ je entropie náhodné veličiny X , $H(Y)$ je entropie náhodné veličiny Y a $H(X,Y)$ je sdružená entropie náhodných veličin X a Y [2]. Na obrázku 9.2 můžeme vidět grafické znázornění vzájemné informace.



Obr. 9.2: Grafické znázornění vzájemné informace

10 VÝSLEDKY

Tato kapitola se věnuje popisu dosažených výsledků. Obsahuje popis dvou způsobů vyhodnocení vypočítaných parametrů obou vybraných úseků. Nejprve jsou parametry analyzovány individuálně, tedy každý zvlášť. Poté jsou pomocí metody SFFS nalezeny takové kombinace parametrů, které vedou ke zlepšení výsledků klasifikace.

10.1 Individuální analýza parametrů

Ze všech 340 vypočítaných parametrů bylo vybráno 10, které vykazovaly největší hodnotu ACC (přesnosti) klasifikace. Ty jsou uvedeny v tabulkách 10.1 a 10.2 pro oba analyzované úseky. Parametry v tabulkách jsou seřazeny sestupně dle hodnoty ACC, které se pro první úsek pohybují v rozmezí hodnot 73,21–66,02 % a pro úsek druhý 60,71–53,57 %. Již z toho faktu je evidentní, že hodnoty přesnosti klasifikace druhého úseku jsou bídné. Jinými slovy druhý úsek není vhodný pro tento typ analýzy, jelikož klasifikace založená na parametrech z toho úseku by byla spíše nahodilého charakteru. Udávaná hranice ACC pro v praxi použitelný vyhodnocovací systém je přes 90 %.

Tab. 10.1: Tabulka výsledků individuální analýzy všech parametrů **prvního** úseku

Parametr	ACC [%]	MI [-]	ρ [-]	Skupina par.
bright 3 k (mean)	73,21	0,5307	0,6475	první
bright 3 k (median)	71,43	0,5165	-0,0953	první
roll of 85 (median)	69,64	0,4701	0,4161	první
bright 3 k (Q3)	69,64	0,4648	0,4183	první
roll of 95 (median)	67,86	0,5844	0,3254	první
roll of 85 (Q3)	67,86	0,4382	0,0616	první
spectral skewness (ir)	67,86	0,4897	0,1827	druhá
spectral centroid (Q3)	67,86	0,4095	-0,0813	druhá
bright 1,5 k (mean)	66,07	0,4935	0,0323	první
spectral flatness (Q3)	66,07	0,4530	-0,0135	druhá

ACC – přesnost, MI – vzájemná informace, ρ – Spearmanův koeficient pořadové korelace, Skupina par. - skupina do, které daný parametr patří viz kapitola 9.3.

Jak můžeme vidět v tabulce 10.1 pro první úsek, nejlepší hodnoty ACC vykazují parametry spadající do prvních dvou skupin parametrů (viz kapitola 9.3). Ty

popisují určitý způsobem barvu zvuku neboli spektrální rozložení. Největší hodnotu přesnosti klasifikace $ACC = 73,21\%$ a zároveň i hodnotu Spearmanova koeficientu pořadové korelace $\rho = 0,6475$ vykazuje parametr bright 3k (mean). Největší hodnotu vzájemné informace $MI=0,5844$ nese parametr roll of 95 (median).

Tab. 10.2: Tabulka výsledků individuální analýzy všech parametrů **druhého** úseku

Parametr	ACC [%]	MI [-]	ρ [-]	Skupina par.
bright 1 k (ir)	60,71	0,3246	0,2967	první
bright 1,5 k (99p)	58,93	0,2396	-0,0402	první
attack slope (ir)	58,93	0,2144	-0,2043	třetí
attack time (iqr)	58,93	0,2007	0,3889	třetí
mfcc7 (var)	57,14	0,2516	0,4256	čtvrtá
mfcc7 (std)	57,14	0,2516	0,4256	čtvrtá
attack slope (99p)	57,14	0,1773	-0,1930	třetí
spectral kurtosis (mean)	55,36	0,1530	0,1129	druhá
zrc (Q3)	53,57	0,1694	-0,3699	pátá
spectral roughness (iqr)	53,57	0,2051	0,1142	pátá

ACC – přesnost, MI – vzájemná informace, ρ – Spearmanův koeficient pořadové korelace, Skupina par. - skupina do, které daný parametr patří viz kapitola 9.3.

V tabulce 10.2 parametrů druhého úseku, stojí za zmínku snad jen parametr bright 1k (ir), který vykazuje nejvyšší hodnotu přesnosti ($ACC = 60,71\%$) a zároveň největší hodnotu vzájemné informace ($MI = 0,3246$). Jak již bylo zmíněno ostatní hodnoty parametrů jsou velmi nízké a pro funkční vyhodnocovací systém prakticky nepoužitelné.

Z individuální analýzy vyplývá, že parametry vypočítané z druhého analyzovaného úseku vykazují nízkou přesnost klasifikace. Tento fakt dávám za vinu jednoduše nevhodně vybranému úseku. Jelikož je tento úsek prakticky po celou dobu hrán ve stejné dynamice (*forte fortissimo*) a s plným nasazením všech nástrojů symfonického orchestru - nedějí se zde žádné velké změny, které by mohly vést k dobrým výsledkům klasifikace. Jsem toho názoru, že i v případě obsáhlejší databáze zvukových nahrávek by výsledek analýzy toho úseku dopadl více méně stejně. Oproti tomu parametry vypočítané z úseku prvního, vykazují mnohem lepší schopnost klasifikovat jednotlivé nahrávky do správných kategorií. Fakt, že nejlepších výsledků dosáhly parametry popisující barvu zvuku by mohl mít vypovídající hodnotu o rozdílných kvalitách záznamu v daných časových kategoriích. Přesto je nutné podotknout, že pro systém využitelný v praxi jsou tyto výsledky stále neuspokojivé.

10.2 Kombinace parametrů

Pomocí metody SFFS byly vypočteny kombinace parametrů, které vedou ke zlepšení úspěšnosti klasifikace. Tato metoda používá kombinaci klasifikace náhodných stromů s validací leave-one-out. Takto byly analyzovány oba vybrané úseky zvlášť a také jejich kombinace.

V tabulce 10.3 můžeme vidět výsledky analýzy prvního úseku. Individuální analýzou bylo dosaženo úspěšnosti klasifikace prvního úseku, pomocí parametru bright 3k (mean), hodnoty ACC=73,21%. Kombinací parametrů bright 3k (mean), spectral roughness (var) a bright 1,5k (Q3) došlo ke zlepšení přesnosti klasifikace o téměř 11%. Dosažená přesnost klasifikace, pomocí kombinace těchto tří parametrů činí ACC = 83,93%.

Tab. 10.3: Tabulka parametrů vybraných metodou SFFS pro analyzovaný první úsek

Pořadí	Parametr	Kombinace	ACC [%]	MI [-]	ρ [-]
1	bright 3k (mean)	1	73,21	0,5307	0,6475
2	spectral roughness (var)	1,2	82,14	0.7409	0.7117
3	bright 1,5k (Q3)	1,2,3	83,93	0.8291	0.7694

Kombinace - kombinace parametrů, ACC – přesnost, MI – vzájemná informace, ρ – Spearmanův koeficient pořadové korelace.

Výsledky analýzy druhého úseku můžeme nalézt v tabulce 10.4. Zde bylo dosaženo zlepšení klasifikace o pouhých 6%, z původní hodnoty přesnosti ACC = 60,71%, na hodnotu ACC = 66,07%. Opět bylo toho výsledku dosaženo kombinací tří parametrů. Tento výsledek není uspokojivý. Nutno ale podotknout, že zázraky se nedějí a již při individuální analýze vypočítaných parametrů druhého úseku bylo konstatováno, že tento úsek není vhodně vybraný, respektive nemá potenciál pro dobré výsledky klasifikace.

Tab. 10.4: Tabulka parametrů vybraných metodou SFFS pro analyzovaný druhý úsek

Pořadí	Parametr	Kombinace	ACC [%]	MI [-]	ρ [-]
1	bright 1k (ir)	1	60,71	0.3246	0.3988
2	attack time (iqr)	1,2	64,29	0.4477	0.3144
3	attack slope (Q1)	1,2,3	66,07	0.4425	0.3984

Kombinace - kombinace parametrů, ACC – přesnost, MI – vzájemná informace, ρ – Spearmanův koeficient pořadové korelace.

V tabulce 10.5 jsou uvedeny výsledky analýzy obou vybraných úseků. Tímto způsobem bylo dosaženo hodnoty přesnosti klasifikace $ACC = 70,54\%$. Tato hodnota není žádným překvapením, když uvážíme, že došlo ke sloučení dvou sad hodnot s rozdílným potenciálem. Jinak řečeno, došlo ke sloučení hodnot úseku prvního, který vykazuje poměrně vysoký potenciál pro dosažení vysokých hodnot klasifikace a hodnot úseku druhého, který vykazuje naopak potenciál mizivý.

Tab. 10.5: Tabulka parametrů vybraných metodou SFFS pro oba analyzované úseky

Pořadí	Parametr	Kombinace	ACC [%]	MI [-]	ρ [-]
1	roll of 85 (Q3)	1	56,25	0.3118	0.1015
2	spectral flatness (99p)	1,2	63,39	0.3288	0.3641
3	mfcc9 (median)	1,2,3	65,18	0.3653	0.3776
4	mfcc9 (Q3)	1,2,3,4	66,07	0.3915	0.4284
5	mfcc12 (median)	1,2,3,4,5	67,86	0.4648	0.4436
6	mfcc2 (1p)	1,2,3,4,5,6	70,54	0.5067	0.4776

Kombinace - kombinace parametrů, ACC – přesnost, MI – vzájemná informace, ρ – Spearmanův koeficient pořadové korelace.

Dosažené výsledky a poznatky implementovaného vyhodnocovacího systému by se daly shrnout do těchto bodů:

- nejvyšší hodnotou přesnosti klasifikace je hodnota $ACC = 83,93\%$. Tato hodnota byla získána kombinací tří parametrů pomocí metody SFFS. Tyto parametry byly získány z prvního analyzovaného úseku (viz tabulka 10.3).
- Nejlepší výsledky přesnosti klasifikace vykazovaly parametry popisující spektrální rozložení energie nebo rozložení spektrálních složek ve spektru.
- Z dosažených výsledků je patrné, že pro tento druh analýzy jsou mnohem více vhodnější úseky s proměnlivou dynamikou a chudším spektrem. Tedy úseky, ve kterých nehraje příliš mnoho hudebních nástrojů.
- Je evidentní, že vytvořená databáze interpretací není natolik obsáhlá (při zachování počtu klasifikačních tříd), aby šlo vytvořit systém použitelný v praxi. Hranice úspěšnosti klasifikace vyhodnocovacího systému použitelného v praxi musí přesáhnout hranici 90%.

11 ZÁVĚR

V první kapitole této diplomové práce je popsána historie a motivace pro vytvoření poměrně mladé interdisciplinární vědecké oblasti MIR.

V následující kapitole jsou nastíněny reprezentace jakých hudba může nabývat. Jsou zde uvedeny zásady pro psaní notového zápisu a jeho jednotlivé elementy. Jsou zde také uvedeny elementární parametry zvukového vlnění.

V kapitole č. 3 je naznačen řetězec, kterým lze hudbu nahrávat a vytvořit jím hudební záznam. Podkapitola 3.1 je věnována zevrubnému popisu A/D převodníku (zvuková karta) a jsou zde popsány základní procesy a jejich parametry, které jsou nezbytné pro A/D převod. Dále jsou uvedeny příklady formáty souborů, které se prakticky užívají pro uložení hudebního záznamu.

Ve čtvrté kapitole, nesoucí název „Základní parametry hudby“ byly shrnuty poznatky čtyř základních parametrů, pomocí kterých můžeme jistým způsobem parametrizovat hudbu. Na tyto parametry je nahlíženo ze dvou pohledů: hudebního a technického. Hudební pohled popisuje jakou funkci daný parametr tvoří reálně v hudbě. Naopak pohled technický přisuzuje danému parametru fyzikální veličinu či princip, pomocí kterého lze parametr popsat.

Následující kapitola je věnována popisu synchronizace hudebních nahrávek. Tento krok je nezbytný pro objektivní porovnání stejných úseků dvou rozdílných interpretací téže skladby.

Kapitola šestá je věnována popisu parametrů, které mohou být extrahovány z jakékoliv hudební nahrávky. Nejprve jsou popsány parametry nízkourovňové. Tyto parametry jsou extrahovány přímo z reprezentací hudebního signálu. Dále jsou uvedeny parametry pomocí, kterých lze popsat barvu zvuku, dynamiku nebo tempo hudební nahrávky. Jako poslední jsou uvedeny vysokoúrovňové statistické parametry.

Dále je uvedena kapitola „Databáze interpretací“ - 8, ve které je popsán rozbor skladby od celosvětově známého klasicistního skladatele Ludwiga von Beethovena, konkrétně první věta - *Allegro con brio* jeho symfonie č. 5, označována názvem „osudová“. Tato databáze je tvořena 56 verzemi této skladby. Jednotlivé hudební nahrávky jsou odlišné interpretace symfonických orchestrů z celého světa. Tyto hudební nahrávky byly pořízeny v různých časových obdobích a jsou proto roztříděny do 3 časových kategorií. Seznam jednotlivých prvků databáze je uveden v příloze B. V podkapitole „Výběr úseků k analýze“ - 8.3 je popsáno jaké úseky byly vybrány k analýze a proč. Cílem bylo vybrat dva co možná nejvíce kontrastní úseky. První úsek má proměnlivou dynamiku a také v tomto úseku nezní celý symfonický orchestr na jednou. Naproti tomu úsek druhý je celý hrán v dynamice *forte fortissimo* a zní zde celý symfonický orchestr na jednou. Jejich notové zápisy nalezneme v přílohách C a D.

Nyní následuje kapitola s názvem „Vyhodnocovací systém“. Zde jsou popsány jednotlivé bloky implementovaného vyhodnocovacího systému, jehož cílem je vyhodnocovat období vzniku zvukových nahrávek do tří klasifikačních kategorií. Tento systém je implementován v prostředí programu MATLAB.

Jako vstupní data pro tento systém slouží vytvořená databáze nahrávek. Nejprve byla provedena synchronizace všech nahrávek databáze pomocí algoritmu DTW, aby bylo možné vyjmout z každé nahrávky dva totožné úseky, které budou podrobeny analýze. Z každého úseku bylo vypočítáno přesně 340 parametrů. Seznam těchto parametrů je uveden v kapitole 9.3. Každý takto vypočítaný parametr má formu vektoru, proto byly z těchto parametrů vypočítány vysokoúrovňové statistické parametry. Tento postup vedl k vytvoření dvou matic parametrů o rozměrech 56x340.

Obě matice parametrů byly podrobeny individuální analýze. Dosažené nejlepší výsledky přesnosti klasifikace se pro první analyzovaný úsek pohybují v rozmezí $ACC = 73,21\text{--}66,02\%$ a pro úsek druhý $ACC = 60,71\text{--}53,57\%$. Z těchto hodnot je evidentní, že druhý úsek nevykazuje potenciál pro tento druh analýzy. Dále bylo zjištěno, že největší schopnost rozlišovat období vzniku interpretace vykazují parametry popisující rozložení spektrální energie a parametry popisující rozložení spektrálních složek. Tento fakt má vypovídající hodnotu o rozdílných kvalitách záznamu v daných časových kategoriích.

Dále byly vypočítané parametry podrobeny analýze metodou SFFS. Tato metoda slouží pro selekci parametrů a zároveň hledá kombinace takových parametrů, které vedou ke zlepšení úspěšnosti klasifikace. Pomocí této metody byly vybrány kombinace tří parametrů a bylo dosaženo úspěšnosti klasifikace prvního úseku $ACC = 83,93\%$. Došlo tedy ke zlepšení přesnosti klasifikace o 11%. Přestože úspěšnost klasifikace $ACC = 83,93\%$ není špatný výsledek, stále se nejedná o takovou hodnotu, aby bylo možné tento vyhodnocovací systém použít v praxi. Minimální hranice úspěšnosti vyhodnocovacího systému využitelného v praxi je 90%.

Implementovaný vyhodnocovací systém prokázal jistý potenciál, ovšem dosažené výsledky nejsou tak dobré, aby bylo možné ho využít v praxi. Zlepšení výsledků by bylo možné dosáhnout rozšířením databáze interpretací.

LITERATURA

- [1] BUDÍKOVÁ, M., KRÁLOVÁ M., MAROŠ, M. *Průvodce základními statistickými metodami*. 1. vyd. Praha: Grada, 2010, 272 s. Expert (Grada). ISBN 978-80-247-3243-5.
- [2] COVER, T., THOMAS, J. *Elements of information theory*. New York: John Wiley&Sons, c1991, 542 s. Wiley series in telecommunications. ISBN 0471062596.
- [3] ČERNOCKÝ, J. *Zpracování řečových signálů*. Brno, 2006. Dostupné také z: <<http://bit.ly/2kskxgM>>. Studijní opora. Vysoké učení technické v Brně, Fakulta informačních technologií, Ústav počítačové grafiky a multimédií.
- [4] ČSN 01 1600: *Akustika - Terminologie*. Česká technická norma, Český normalizační institut, únor 2003.
- [5] DOWNIE, J. S. Music information retrieval. *Annual Review of Information Science and Technology* [online]. 2003, **37**(1), 295-340 [cit. 2017-12-1]. DOI: 10.1002/aris.1440370108. ISSN 00664200. Dostupné z: <<http://doi.wiley.com/10.1002/aris.1440370108>>
- [6] FELLER, W. *An introduction to probability theory and its applications*. 3rd ed. New York: John Wiley, c1968. Wiley series in probability and mathematical statistics. ISBN 0471257087.
- [7] KLASCHKA, J., KOTRČ, E. *Klasifikační a regresní lesy*. ROBUST 2004. Sborník prací 13. letní školy JČMF, 2004.
- [8] KNEES, P., SCHEDL, M. *Music similarity and retrieval: an introduction to audio- and web-based strategies*. Vol. 36. New York, NY: Springer Berlin Heidelberg, 2016. ISBN 9783662497203.
- [9] KOHAVI, R., JOHN, G. H. *Wrappers for feature subset selection*. *Artificial Intelligence*, 1997, č. 1–2, s. 273-324.
- [10] LARTILLOT, O. MIRtoolbox 1.7 User's Manual [online]. 2017 [cit. 2017-11-11]. Dostupné z URL: <<https://www.jyu.fi/hytk/fi/laitokset/mutku/en/research/materials/mirtoolbox/manual1-7.pdf>>.
- [11] MUSICBRAINZ, Symphony no. 5 in C minor, op. 67: I. Allegro con brio [online]. 2014 [cit. 2018-03-07]. Dostupné z URL: <<https://musicbrainz.org/work/b6f9ecc3-24d1-38ed-b8a0-091f7cd0c6b2>>.

- [12] MEKYSKA J., GALÁŽ Z., MŽOUREK Z., SMÉKAL Z. *Sequential floating feature selection tool* [online]. [cit. 2018-03-16]. Dostupné z URL: <<http://splab.cz/download/software/software-pro-sekvencni-selekcii-priznaku>>.
- [13] MERHAUT, J. a kol. *Průručka elektroakustiky*. Státní nakladatelství technické literatury, Praha 1964. 608 s. 04-550-64
- [14] MÜLLER, M. *Fundamentals of music processing*. 1. New York, NY: Springer Berlin Heidelberg, 2015. ISBN 9783319219448.
- [15] NEUŽIL, J., BENKO, M. *Skripta základů hudební teorie*. Praha, 2010. Dostupné také z: <http://download.kjj.cz/pub/vyuka/2010/kl/skripta_kl.pdf>. Skriptum. Konzervatoř Jaroslava Ježka.
- [16] PEETERS, G. *A large set of audio features for sound description (similarity and classification) in the CUIDADO project* 2004. [online]. [cit. 2017-12-12]. Dostupné z URL: <http://recherche.ircam.fr/anasyn/peeters/ARTICLES/Peeters_2003_cuidadoaudiofeatures.pdf>.
- [17] PEETERS, G. *The Timbre Toolbox: Extracting audio descriptors from musical signals* 2011. [online]. [cit. 2017-12-07]. Dostupné z URL: <https://www.mcgill.ca/mpcl/files/mpcl/peeters_2011_jasa.pdf>.
- [18] PENG, H., LONG, F., DING, C. *Feature Selection Based on Mutual Information: Criteria of Max-dependency, Max-relevance, and Min-redundancy*. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, ročník 27, č. 8, s. 1226–1238, Aug. 2005.
- [19] PUDIL, P., NOVOVIČOVÁ, J., KITTLER, J. *Floating search methods in feature selection*. *Pattern Recogn. Lett.* 15, 11 (November 1994), 1119-1125.
- [20] RANDEL, D. M. *The Harvard dictionary of music*. 4th ed. Cambridge, Mass.: Belknap Press of Harvard University Press, 2003. ISBN 0674011635.
- [21] SETHARES, W. A. *Tuning, timbre, spectrum, scale*. New York: Springer, c1998. ISBN 9783540761730.
- [22] SENIN, P. *Dynamic time wrapping algorithm review*. Honolulu, USA, 2008. Dostupné také z: <<http://bit.ly/2BokPQm>>. Skriptum. University of Hawaii at Manoa, Information and Computer Science Department.
- [23] SCHEDL, M., GOMEZ, E., URBANO, J. *Music Information Retrieval: Recent Developments and Applications*. Vol. 8. Boston, USA: Now Publishers, 2014. ISBN 978-1601988065.

- [24] SCHIMMEL, J. *Elektroakustika*. Brno, 2014. Skriptum. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací.
- [25] SCHIMMEL, J. *Studiová a hudební elektronika*. Brno, 2015. Skriptum. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací.
- [26] SMÉKAL, Z. *Číslíkové zpracování řeči*. Brno, 2010. Skriptum. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací.
- [27] TOBIAS, J. V. *Foundations of modern auditory theory*. Vol. 2. New York [u.a.]: Academic Press, 1972. ISBN 9780126919028.
- [28] TZANETAKIS, G. *Music Information Retrieval*: University of Victoria, Department of Computer Science, 2015. Dostupné z URL: <http://www.marsyas.info/mirBook/mirBook_jan05_2015.pdf>.
- [29] WEIHS, C., et al. (ed.). *Music Data Analysis: Foundations and Applications*. CRC Press, 2016.

SEZNAM SYMBOLŮ, VELIČIN A ZKRATEK

ACC	<i>Accuracy</i> - přesnost
A/D	<i>Analogově digitální převodník</i>
BER	<i>Band Energy Ratio</i>
BMP	<i>Beats per minute</i> - počet úderů na minutu
DTW	<i>Dynamic time wrapping</i>
FFT	<i>Fast Fourier Transform</i> - rychlá Fourierova transformace
mRMR	<i>minimum Redundancy Maximum Relevance</i> - metoda minimální redundance, maximální relevance
IR	<i>Interpercentile range</i> - mezikpercentilové rozpětí
IQR	<i>Interquartile range</i> - mezikvartilové rozpětí
MIR	<i>Music information retrieval</i>
NSF	<i>National Science Foundation</i>
RMS	<i>Root Mean Square</i>
PCM	<i>Pulse Code Modulation</i>
SC	<i>Spectral Centroid</i>
SF	<i>Spectral Flux</i>
SFFS	<i>Sequential Forward Floating Search</i>
ZRC	<i>Zero crossing rate</i>

SEZNAM PŘÍLOH

A	Obsah přiloženého CD	65
B	Databáze nahrávek	66
C	Notový zápis prvního úseku	68
D	Notový zápis druhého úseku	70

A OBSAH PŘILOŽENÉHO CD

Na přiloženém CD se nachází:

- diplomová práce ve formátu pdf,
- složka se zdrojovými soubory `vyhodnocovací_system`,
- kompletní notový zápis analyzované skladby ve formátu pdf.

Původní wav soubory zvukových nahrávek databáze, nejsou součástí přiloženého CD z důvodů jejich velikosti.

B DATABÁZE NAHRÁVEK

Tab. B.1: Tabulka nahrávek v databázi

Číslo	Dirigent	Název orchestru	Rok nahrání
1	A. Cluytens	The Berlin Philharmonic Orchestra	1958
2	A. Doráti	London Symphony Orchestra	1962
3	A. Nanut	Radio Symphony Orchestra Ljubljana	1986
4	A. Polizz	Budapest Symphony Orchestra	2002
5	A. Toscanini	NBC Symphony Orchestra	1952
6	B. Drahos	Nicolaus Esterházy Sinfonia	1997
7	B. Haitink	London Symphony Orches	2006
8	B. Güller	Bamberger Symphoniker	2005
9	C. M. Giulini	Los Angeles Philharmonic	1983
10	C. Kleiber	Wiener Philharmoniker	1974
11	C. Païta	Philharmonic Symphony Orchestra	1981
12	C. Abbado	Wiener Philharmoniker	1987
13	D. Barenboim	Staatskapelle Berlin	1999
14	D. Barenboim	West-Eastern Divan Orchestra	2012
15	D. Parry	London Philharmonic Orchestra	2005
16	E. Ormandy	The Philadelphia Orchestra	1966
17	E. Bilyar	Armonie Symphony Orchestra	2010
18	F. Fricsay	Berliner Philharmoniker	1961
19	F. Konwitschny	Gewandhausorchester Leipzig	1961
20	F. Reiner	Chicago Symphony Orchestra	1990
21	F. Singer	National Symphony Orchestra	1988
22	G. Szell	Cleveland Orchestra	1963
23	G. Cantelli	NBC Symphony Orchestra	1950
24	G. Wand	NDR Sinfonieorchester	1987
25	H. Hartmann	Orchestra Magdeburg Symphony	2010
26	H. Blomstedt	Staatskapelle Dresden	1977
27	H. von Karajan	Berliner Philharmoniker	1982
28	H. von Karajan	Philharmonia Orchestra	1954
29	H. von Karajan	Wiener Philharmoniker	1948
30	J. Ferencsik	Hungarian State Orchestra	1975
31	J. Lamon	Tafelmusik Orchestra	2004
32	J. E. Gardiner	Orchestre Rév. et Romantique	1994
33	J. Krips	London Symphony Orchestra	1960
34	J. Keilberth	Hamburg Philharmonic Orchestra	1959

Tab. B.1: Tabulka nahrávek v databázi

Číslo	Dirigent	Název orchestru	Rok nahrání
35	K. Ancerl	Czech Philharmonic Orchestra	1953
36	K. Böhm	Wiener Philharmoniker	1977
37	K. Masur	New York Philharmonic	1993
38	K. Sanderling	Philharmonia Orchestra	1981
39	L. Bernstein	New York Philharmonic	1961
40	M. Bittner	Moscow Philharmonic Orchestra	2009
41	N. Harnoncourt	Chamber Orchestra of Europe	1990
42	O. Klemperer	Philharmonia Orchestra	1960
43	P. Henry Smith	The Fauxharmonic Orchestra	2016
44	P. Eotvos	Göteborgs Symfoniker	2001
45	R. F. de Burgos	London Symphony Orchestra	1990
46	R. Kubelik	Boston Symphony Orchestra	1973
47	R. Edlinger	Zagreb Philharmonic Orchestra	1988
48	R. Kempe	Münchner Philharmoniker	1971
49	S. Ozawa	Chicago Symphony Orchestra	1970
50	Sir A. Boult	Promenade Orchestra of London	1966
51	Sir G. Solti	Chicago Symphony Orchestra	1975
52	Sir G. Solti	Wiener Philharmoniker	1959
53	Sir J. Barbirolli	Hallé Orchestra	1958
54	Sir R. Norrington	London Classical Players	1990
55	V. Neumann	Czech Philharmonic Orchestra	1990
56	W. Furtwängler	Berliner Philharmoniker	1976

C NOTOVÝ ZÁPIS PRVNÍHO ÚSEKU

Fl. A

Ob.

Cl.

Fag.

Cor.

Tr.

Tp.

ff *ten.* *len.* *p*

pizz. *p* *pizz.* *p* *pizz.* *p*

p

Fl.

Ob.

Cl.

Fag.

Tr.

Tp.

cresc. *Adagio.* *f* *p*

cresc. *cresc.* *f* *p*

cresc. *cresc.* *f* *p*

cresc. *arco* *cresc.* *ff* *arco.* *p* *cresc.*

Bassi. *cresc.* *ff* *p* *cresc.*

Fl.

Ob.

Cl.

Fag.

Cor.

Tr.

Tp.

D NOTOVÝ ZÁPIS DRUHÉHO ÚSEKU

Fl.
Ob.
Cl.
Fag.
Cor.
Tr.
Tp.

Fl.
Ob.
Cl.
Fag.
Cor.
Tr.
Tp.

A

70

Fl.

Ob.

Cl.

Fag.

Cor.

Tr.

Tp.

Vcl.

Basso.

f

Fl.

Ob.

Cl.

Fag.

Cor.

Tr.

Tp.

Bassi.

f

Fl.

Ob.

Cl.

Fag.

Cor.

Tr.

Tp.

Basso.

cel.

ff

ff

ff

Basso.

cel.

Fl.

Ob.

Cl.

Fag.

Cor.

Tr.

Tp.

B

ff

ff

ff

B